

\*\*\*\*\*

MSRCH\_PP protein - protein database search, using Smith-Waterman algorithm

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

On: Mon Jul 12 12:08:57 1999; MasPar time 11.79 Seconds  
775.113 Million cell updates/sec  
Regular output not generated.

Title: >US-08-911-423-2  
Description: (1-228) from US08911423.pap  
Perfect Score: 228  
Sequence: 1 MGVAMLYGVSMCLVLDLGO.....PEERGRQTEKCHLGRWP 228

Scoring table: TABLE uniprotatable  
Gap 60

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 3.330; Variance 0.398; scale 8.368

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	7	3.1	100	B64472	hypothetical protein	3.58e-01
2	7	3.1	218	B64365	hypothetical protein	3.58e-01
3	7	3.1	365	1 MNXRAH	nonstructural protein	3.58e-01
4	7	3.1	534	S21961	proline-rich protein	3.58e-01
5	7	3.1	575	A69716	spore coat assembly p	3.58e-01
6	7	3.1	663	S69626	hypothetical protein	3.58e-01
7	7	3.1	773	A60011	probable pre-mRNA-spl	3.58e-01
8	7	3.1	791	E71333	hypothetical protein	3.58e-01
9	6	2.6	19	C37072	proteoglycan 65K core	2.80e-01
10	6	2.6	85	A05180	cytochrome c6 - Sync	2.80e-01
11	6	2.6	89	I49515	B14 protein B - mous	2.80e-01
12	6	2.6	97	S60846	M protein precursor -	2.80e-01
13	6	2.6	111	B37321	tRNA (uracil-5-)meth	2.80e-01
14	6	2.6	111	J01083	cytochrome c6 precurs	2.80e-01
15	6	2.6	120	B25429	T-cell receptor beta	2.80e-01
16	6	2.6	120	B69898	hypothetical protein	2.80e-01
17	6	2.6	135	S44214	genome polypeptide -	2.80e-01
18	6	2.6	140	H71903	hypothetical protein	2.80e-01
19	6	2.6	141	I40885	cytotoxin - Clostridi	2.80e-01
20	6	2.6	160	1 AFMWA	allophycocyanin alpha	2.80e-01
21	6	2.6	161	F71117	hypothetical protein	2.80e-01
22	6	2.6	161	S61389	small basic protein s	2.80e-01
23	6	2.6	161	S74735	allophycocyanin-B - S	2.80e-01

24	2.6	161	S00870	allophycocyanin B alp	2.80e-01
25	2.6	176	S15913	hypothetical protein	2.80e-01
26	2.6	176	T03420	traf protein - Agroba	2.80e-01
27	2.6	180	A61433	trypsin inhibitor 2a	2.80e-01
28	2.6	182	S40729	hypothetical protein	2.80e-01
29	2.6	189	S60587	non-structural protei	2.80e-01
30	2.6	199	C64949	yecD protein - Escher	2.80e-01
31	2.6	201	S51424	hypothetical protein	2.80e-01
32	2.6	213	T01464	hypothetical protein	2.80e-01
33	2.6	226	JN0464	insulin-like growth f	2.80e-01
34	2.6	229	D35116	hypothetical protein	2.80e-01
35	2.6	236	B70488	cytochrome-c oxidase	2.80e-01
36	2.6	238	I48605	insulin-like growth f	2.80e-01
37	2.6	239	S60083	conserved hypotherica	2.80e-01
38	2.6	246	S27963	modulator recognition	2.80e-01
39	2.6	249	B70440	conserved hypotherica	2.80e-01
40	2.6	253	D69498	peptidyl-prolyl cis-t	2.80e-01
41	2.6	260	B64563	hypothetical protein	2.80e-01
42	2.6	281	D74426	phosphate transport s	2.80e-01
43	2.6	287	S71192	mitosis-specific cycl	2.80e-01
44	2.6	299	C70843	hypothetical protein	2.80e-01
45	2.6	333	J50590	endo-1,4-beta-xylanas	2.80e-01
46	2.6	343	C69212	conserved hypotherica	2.80e-01
47	2.6	356	S56417	hypothetical 40.3K pr	2.80e-01
48	2.6	357	F69987	spore coat protein ho	2.80e-01
49	2.6	366	A37321	tRNA (uracil-5-)meth	2.80e-01
50	2.6	369	S74017	hypothetical protein	2.80e-01
51	2.6	381	S60623	ubiquinol--cytochrome	2.80e-01
52	2.6	392	B69321	cell division protein	2.80e-01
53	2.6	393	C71254	hypothetical protein	2.80e-01
54	2.6	393	B44767	L-mandelate dehydroge	2.80e-01
55	2.6	401	S76788	histidine--trna ligas	2.80e-01
56	2.6	414	B71125	probable cell divisio	2.80e-01
57	2.6	443	S37612	NADH dehydrogenase (u	2.80e-01
58	2.6	443	1 DEBPA	pyruvate dehydrogenas	2.80e-01
59	2.6	463	A49898	regulatory gene 5' of	2.80e-01
60	2.6	465	S46759	hypothetical protein	2.80e-01
61	2.6	465	S76464	hypothetical protein	2.80e-01
62	2.6	489	S76768	hypothetical protein	2.80e-01
63	2.6	493	T03609	cyclin, A-type - comm	2.80e-01
64	2.6	532	T02721	probable methylmalona	2.80e-01
65	2.6	537	1 Q0E30	BRP2 protein - human	2.80e-01
66	2.6	548	JC5090	pyoverdine synthetase	2.80e-01
67	2.6	555	E70409	carbon starvation pro	2.80e-01
68	2.6	574	A69196	cell surface glycopro	2.80e-01
69	2.6	575	S59740	NRD1 protein - yeast	2.80e-01
70	2.6	587	S58319	hypothetical protein	2.80e-01
71	2.6	643	T03631	oxygenase, pathogen-i	2.80e-01
72	2.6	660	H71423	hypothetical protein	2.80e-01
73	2.6	723	T02477	hypothetical protein	2.80e-01
74	2.6	729	S68191	triadin - human	2.80e-01
75	2.6	746	S31816	genome polypeptide -	2.80e-01
76	2.6	758	D71072	hypothetical protein	2.80e-01
77	2.6	829	I46536	Ksp-cadherin - rabbit	2.80e-01
78	2.6	870	1 XPBE12	major antigenic struc	2.80e-01
79	2.6	932	1 RBHU	retinoblastoma-associ	2.80e-01
80	2.6	932	S65214	probable alpha/gamma	2.80e-01
81	2.6	936	I40705	bacterial adhesin - C	2.80e-01
82	2.6	951	T03511	ice nucleation protel	2.80e-01
83	2.6	986	S12021	thermoactive cellulas	2.80e-01
84	2.6	1030	T00812	chloroplast envelope	2.80e-01
85	2.6	1180	1 NCEKX5	exodeoxyribonuclease	2.80e-01
86	2.6	1201	A57369	anillin - fruit fly (	2.80e-01
87	2.6	1274	I40487	surfactin synthetase	2.80e-01
88	2.6	1433	G01946	nitric-oxide synthase	2.80e-01
89	2.6	1483	S42839	Ti6G12.5 protein - Ca	2.80e-01
90	2.6	1876	S50235	1,3-beta-glucan synth	2.80e-01
91	2.6	1895	S50240	1,3-beta-glucan synth	2.80e-01
92	2.6	2227	1 GNNYMK	genome polypeptide -	2.80e-01
93	2.6	2227	1 GNNYHM	genome polypeptide -	2.80e-01
94	2.6	3010	S18030	genome polypeptide -	2.80e-01
95	2.6	3010	A45573	polypeptide(protein N	2.80e-01
96	2.6	3010	1 GNVWCJ	genome polypeptide -	2.80e-01

97 6 2.6 3011 2 S40770 polyprotein precursor 2.80e+01  
 98 6 2.6 3011 1 GNWVCH genome polyprotein - 2.80e+01  
 99 6 2.6 3924 2 S7431 ankylin 2, neuronal 1 2.80e+01  
 100 6 2.6 4151 2 G70944 probable polyketide s 2.80e+01

## ALIGNMENTS

RESULT 1  
 ENTRY #type complete  
 TITLE hypothetical protein MJ1379 - Methanococcus jannaschii  
 ORGANISM #formal\_name Methanococcus jannaschii  
 DATE 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 13-Sep-1998  
 B64472  
 A64300  
 Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 Science (1996) 273:1058-1073  
 Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.  
 #cross-references MUID:96337999  
 #accession B64472  
 #status Preliminary; nucleic acid sequence not shown;  
 #molecule\_type DNA  
 #residues 1-100 #label BUL  
 #cross-references GB:U67578; GB:L77117; NID:g1592024; PID:g1592331;  
 TIGR:MJ1379; PID:g1511396

GENETICS  
 #map\_position FOR1328526-1328828  
 CLASSIFICATION #superfamily conserved hypothetical protein MJ0128  
 SUMMARY #length 100 #molecular-weight 11819 #checksum 7922  
 Query Match 3.1%; Score 7; DB 2; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 3.58e-01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 35 RGEOTEE 41  
 1111111  
 213 RGEOTEE 219

RESULT 2  
 ENTRY #type complete  
 TITLE hypothetical protein MJ0522 - Methanococcus jannaschii  
 ORGANISM #formal\_name Methanococcus jannaschii  
 DATE 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Oct-1997  
 B64365  
 A64300  
 Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 Science (1996) 273:1058-1073  
 Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.  
 #cross-references MUID:96337999

#accession B64365 Preliminary; nucleic acid sequence not shown;  
 #status translation not shown  
 #molecule\_type DNA  
 #residues 1-218 #label BUL  
 #cross-references GB:U67502; GB:L77117; NID:g1591223; PID:g1591225;  
 TIGR:MJ0522; PID:g1510596

GENETICS  
 #map\_position REV462424-461768  
 SUMMARY #length 218 #molecular-weight 23896 #checksum 8835  
 Query Match 3.1%; Score 7; DB 2; Length 218;  
 Best Local Similarity 100.0%; Pred. No. 3.58e-01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 16 GDIVFGF 22  
 1111111  
 92 GDIVFGF 98  
 RESULT 3  
 ENTRY #type complete  
 TITLE nonstructural protein NS2 - African horse sickness virus (serotype 9)  
 ORGANISM #formal\_name African horse sickness virus  
 DATE 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 05-Sep-1997  
 A40788  
 A40788  
 #accession A40788  
 #authors van Staden, V.; Theron, J.; Greyling, B.J.; Huismans, H.; Nel, L.H.  
 #journal virology (1991) 185:500-504  
 #title A comparison of the nucleotide sequences of cognate NS2 genes of three different orbiviruses.  
 #cross-references MUID:92024120  
 #accession A40788  
 #molecule\_type genomic RNA  
 #residues 1-365 #label VAN  
 #cross-references GB:M69090; NID:g210058; PID:g210059

GENETICS  
 #map\_position segment 8  
 CLASSIFICATION #superfamily bluetongue virus nonstructural protein NS2  
 KEYWORDS nonstructural protein; RNA binding  
 SUMMARY #length 365 #molecular-weight 41193 #checksum 8178  
 Query Match 3.1%; Score 7; DB 1; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 3.58e-01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 15 CVLDLQG 21  
 1111111  
 14 CVLDLQG 20

RESULT 4  
 ENTRY #type complete  
 TITLE proline-rich protein APG - Arabidopsis thaliana  
 ORGANISM #formal\_name Arabidopsis thaliana #common\_name mouse-ear cress  
 DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Sep-1997  
 S21961  
 S21961  
 S16748  
 #accession S21961  
 #molecule\_type DNA  
 #residues 1-534 #label ROB  
 #cross-references EMBL:X60377; NID:g22598; PID:g22599

```

GENETICS
#gene
#introns
SUMMARY
#length 534 #molecular-weight 57967 #checksum 9743

Query Match 3.1%; Score 7; DB 2; Length 534;
Best Local Similarity 100.0%; Pred. No. 3.58e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 CIFEFLT 32
      |||||
QY 164 CIFEFLT 170

RESULT 5
ENTRY
TITLE
ORGANISM
DATE
#accessions
REFERENCE
#authors

A69716 #type complete
spore coat assembly protein (spoVID) - Bacillus subtilis
#formal_name Bacillus subtilis
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
02-Jul-1998
A69716: B47083
A69580
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Serror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
Y.; Yata, K.; Yoshida, K.; Yoshikawa, H.P.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession A69716
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-575 #label KUN
#cross-references GB:299118; GB:AL009126; NID:g2635200; PID:el184060;
PID:g2635276
#experimental_source strain 168
REFERENCE
A47083
#authors Beall, B.; Driks, A.; Losick, R.; Moran Jr., C.P.
#journal J. Bacteriol. (1993) 175:1705-1716
#title Cloning and characterization of a gene required for assembly
of the Bacillus subtilis spore coat.
#cross-references MUID:93194796
#accession B47083
#status preliminary
#molecule_type nucleic acid
#residues 'MNL', 2-575 #label BEA
#note sequence extracted from NCBI backbone (NCBIN:127855,
NCBIP:127857)

GENETICS
#gene
SUMMARY
#length 575 #molecular-weight 64976 #checksum 5576

Query Match 3.1%; Score 7; DB 2; Length 575;
Best Local Similarity 100.0%; Pred. No. 3.58e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 168 SVVEEPG 174
      |||||
QY 22 SVVEEPG 28

RESULT 6
ENTRY
TITLE
ORGANISM
DATE
#accessions
REFERENCE
#authors
#submission
#description
#accession
#molecule_type DNA
#residues 1-663 #label DIE
#cross-references EMBL:U33050; NID:g927726; PID:g927762; MIPS:YDR458C

S69626 #type complete
hypothetical protein YDR458c - yeast (Saccharomyces
cerevisiae)
#formal_name Saccharomyces cerevisiae
22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change
21-Nov-1997
S69626
S69553
Dietrich, F.S.
#submission submitted to the EMBL Data Library, August 1995
#description The sequence of S. cerevisiae cosmids 9410, 8035, 8166, and
9787.
#accession S69626
#molecule_type DNA
#residues 1-663 #label DIE
#cross-references EMBL:U33050; NID:g927726; PID:g927762; MIPS:YDR458C

GENETICS
#map_position 4R
SUMMARY
#length 663 #molecular-weight 76376 #checksum 2274

Query Match 3.1%; Score 7; DB 2; Length 663;
Best Local Similarity 100.0%; Pred. No. 3.58e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 605 FLTTVQL 611
      |||||
QY 167 FLTTVQL 173

RESULT 7
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE
#accessions
REFERENCE
#authors
#submission
#accession
#molecule_type DNA
#residues 1-773 #label BEC
#cross-references EMBL:Z36011; NID:g536441; PID:g536442; MIPS:YBR142w
S45011 #type complete
probable pre-mRNA-splicing RNA helicase - yeast
(Saccharomyces cerevisiae)
#formal_name Saccharomyces cerevisiae
26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change
05-Jun-1998
S45011; S46013; S46620
S45995
Becam, A.M.; Herbert, C.J.; Nasr, F.; Slonimski, P.P.;
Zagulski, M.
#submission submitted to the Protein Sequence Database, August 1994
#accession S46011
#molecule_type DNA
#residues 1-773 #label BEC
#cross-references EMBL:Z36011; NID:g536441; PID:g536442; MIPS:YBR142w
S46013
Entian, K.D.; Koetter, P.; Rose, M.; Becker, J.; Grey, M.;
Li, Z.; Niemann, E.; Schenk-Groeninger, R.; Servos, J.;
Wehner, E.; Wolter, R.; Brendel, M.; Bauer, J.; Braun, H.;
Dern, K.; Duesterhus, S.; Gruenbein, R.; Hedges, D.;
Kiesau, P.; Korol, S.; Krems, B.; Proft, M.; Siegers, K.;

```

```

Baur, A.; Boles, E.; Miosga, T.; Schaaff-Gerstenschlaeger,
I.; Zimmermann, F.K.
#submission submitted to the Protein Sequence Database, August 1994
#accession S46013
#molecule_type DNA
#residues 770-773 #label ENT
#cross-references EMBL:Z36011; MIPS:YBR142w
#experimental_source strain S288C
REFERENCE S46619
#authors Zagulski, M.; Becam, A.M.; Grzybowska, E.; Lacroute, F.;
Migdalinski, A.; Slonimski, P.P.; Sokolowska, B.; Herbert,
C.J.
#journal Yeast (1994) 10:1227-1234
#title The sequence of 12.5 kb from the right arm of chromosome II
predicts a new N-terminal sequence for the IRA1 protein and
reveals two new genes, one of which is a DEAD-box helicase.
#cross-references MUID:95274325
#accession S46620
#status nucleic acid sequence not shown; not compared with
conceptual translation
#molecule_type DNA
#residues 198-577 #label ZAG
#experimental_source strain S288C
GENETICS
#gene SGD:MAK5
#cross-references SGD:S0000346; MIPS:YBR142w
#map_position 2R
#classification #superfamily unassigned DEAD/H box helicases; DEAD/H box
helicase homology
KEYWORDS ATP; DEAD box; P-loop
FEATURE
215-572 #domain DEAD/H box helicase homology #label DEAD\
215-222 #region nucleotide-binding motif A (P-loop)\
329-334 #region nucleotide-binding motif B\
333-336 #region DEAD motif\
221 #binding_site ATP (lys) #status predicted
SUMMARY #length 773 #molecular-weight 87047 #checksum 2101
Query Match 3.1%; Score 7; DB 2; Length 773;
Best Local Similarity 100.0%; Pred. No. 3.58e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 458 FLTTPPG 464
Qy 127 FLTTPPG 133
ULT 8
ULT #type complete
TITLE hypothetical protein TP0374 - syphilis spirochete
ORGANISM #formal_name treponema pallidum subsp. pallidum #common_name
syphilis spirochete
DATE 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change
17-Mar-1999
ACCESSIONS E71333
REFERENCE A71350
#authors Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.;
Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton,
R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod,
M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson,
D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald,
L.; Artisch, P.; Bowman, C.; Cotton, M.D.; Fujii, C.;
Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey,
L.; Weidman, J.; Smith, H.O.; Venter, J.C.
#journal Science (1998) 281:375-388
#title Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
#cross-references MUID:98332770
#accession E71333
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-791 #label COL

```

```

#cross-references GB:AE001216; GB:AE000520; NID:g3322647; PID:g3322655
#experimental_source strain Nichols
GENETICS
#gene TP0374
SUMMARY #length 791 #molecular-weight 89451 #checksum 180
Query Match 3.1%; Score 7; DB 2; Length 791;
Best Local Similarity 100.0%; Pred. No. 3.58e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 559 EPLPTEQ 565
Qy 144 EPLPTEQ 150
RESULT 9
ENTRY #type fragment
TITLE proteoglycan 65K core peptide KS-C - chicken (fragment)
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 15-Feb-1991 #sequence_revision 15-Feb-1991 #text_change
30-Sep-1993
ACCESSIONS C37072
REFERENCE A37072
#authors Krueger Jr., R.C.; Fields, T.A.; Hildreth IV, J.; Schwartz,
N.B.
#journal J. Biol. Chem. (1990) 265:12075-12087
#title Chick cartilage chondroitin sulfate proteoglycan core
protein. I. Generation and characterization of peptides and
specificity for glycosaminoglycan attachment.
#cross-references MUID:90307743
#accession C37072
#status preliminary
#molecule_type protein
#residues 1-19 #label KRU
SUMMARY #length 19 #checksum 4524
Query Match 2.6%; Score 6; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 2.80e-01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 PSVVEE 8
Qy 21 PSVVEE 26
RESULT 10
ENTRY #type complete
TITLE cytochrome c6 - Synecococcus sp.
ALTERNATE_NAMES cytochrome c553; soluble cytochrome f
ORGANISM #formal_name Synecococcus sp.
DATE 05-Jun-1987 #sequence_revision 13-Sep-1996 #text_change
15-Jan-1999
ACCESSIONS A05180
REFERENCE A94469
#authors Margolias, E.
#citation unpublished results, cited by Dickerson, R.E., in The
Evolution of Protein Structure and Function, Sigman, D.S.,
and Brazier, M.A.B., eds., pp.173-202, Academic Press, New
York and London, 1980
#contents Anacystis nidulans
#accession A05180
#residues 1-85 #label MAR
COMMENT Synecococcus is a genus of blue-green algae.
CLASSIFICATION #superfamily cytochrome c6; cytochrome c6 homology
chromoprotein; electron transfer; heme; iron; photosynthesis
FEATURE
4-77 #domain cytochrome c6 homology #label CYC\
14-17 #binding_site heme (Cys) (covalent) #status predicted\
18,58 #binding_site heme iron (His, Met) (axial ligands)
#status predicted
SUMMARY #length 85 #molecular-weight 8743 #checksum 4549

```

```

Query Match          2.6%; Score 6; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 2.80e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 CHLGR 22
|||||
QY 221 CHLGR 226

RESULT 11
ENTRY B144 protein B - mouse (fragment)
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
DATE 02-Jul-1996
ACCESSIONS I49515 #type fragment
REFERENCE I49514
AUTHORS Tsuge, I.; Shen, F.
JOURNAL Immunogenetics (1987) 26:378-380
TITLE A gene in the H-2S: H-2D interval of the major
histocompatibility complex which is transcribed in B cells
and macrophages.
#cross-references MUID:88031493
#accession I49515
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-89 #label RES
#cross-references GB:M18187; NID:g192097; PID:g192099
SUMMARY #length 89 #checksum 7128

Query Match          2.6%; Score 6; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 2.80e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 NGSGN 22
|||||
QY 36 NGSGN 41

RESULT 12
ENTRY M protein precursor - Streptococcus pyogenes (serotype
TITLE PT4245) (fragment)
ORGANISM #formal_name Streptococcus pyogenes
#variety serotype PT4245
DATE 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change
ACCESSIONS S60846
REFERENCE S60784
AUTHORS Whatmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.;
Kehoe, M.A.
JOURNAL Mol. Microbiol. (1994) 14:619-631
TITLE Non-congruent relationships between variation in emm gene
sequences and the population genetic structure of group A
streptococci.
#cross-references MUID:95198537
#accession S60846
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-97 #label WHA
#cross-references EMBL:U11966; NID:9533609; PID:gl235824
#note the nucleotide sequence was submitted to the EMBL Data
Library, July 1994
CLASSIFICATION #superfamily M5 protein
SUMMARY #length 97 #checksum 5813

Query Match          2.6%; Score 6; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.80e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 50 EERGE 55
|||||
QY 210 EERGE 215

RESULT 13
ENTRY B37321 #type fragment
TITLE tRNA (uracil-5)-methyltransferase (EC 2.1.1.35) - Salmonella
typhimurium (fragment)
ORGANISM #formal_name Salmonella typhimurium
DATE 11-Sep-1992 #sequence_revision 11-Sep-1992 #text_change
23-Jun-1993
ACCESSIONS B37321
REFERENCE A37321
AUTHORS Gustafsson, C.; Lindstroem, P.H.R.; Hagervall, T.G.; Esberg,
K.B.; Bjoerk, G.R.
JOURNAL J. Bacteriol. (1991) 173:1757-1764
TITLE The trna promoter has regulatory features and sequence
elements in common with the rRNA P1 promoter family of
Escherichia coli.
#cross-references MUID:91154132
#accession B37321
#status preliminary
#molecule_type DNA
#residues 1-111 #label GUS
#cross-references GB:M57569
KEYWORDS methyltransferase
SUMMARY #length 111 #checksum 8571

Query Match          2.6%; Score 6; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 2.80e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 LPTEQY 11
|||||
QY 146 LPTEQY 151

RESULT 14
ENTRY JQ1083 #type complete
TITLE cytochrome c6 precursor - Synecococcus sp. (strain PCC 7942)
ALTERNATE_NAMES cytochrome c553; soluble cytochrome f
ORGANISM #formal_name Synecococcus sp.
DATE 31-Dec-1991 #sequence_revision 13-Sep-1996 #text_change
15-Jan-1999
ACCESSIONS JQ1083
REFERENCE JQ1083
AUTHORS Laudénbach, D.E.; Herbert, S.K.; McDowell, C.; Fork, D.C.;
Grossman, A.R.; Straus, N.A.
JOURNAL Plant Cell (1990) 2:913-924
TITLE Cytochrome c-553 is not required for photosynthetic activity
in the cyanobacterium Synecococcus.
#cross-references MUID:93005680
#accession JQ1083
#molecule_type DNA
#residues 1-111 #label LAU
COMMENT This protein functions as a mobile carrier of electrons between the
membrane-bound cytochrome b6-f complex and the p-700 reaction
center of photosystem I in cyanobacteria and many eukaryotic
algae.
GENETICS cyta
#gene
CLASSIFICATION #superfamily cytochrome c6; cytochrome c6 homology
KEYWORDS chromoprotein; electron transfer; heme; iron; photosynthesis
FEATURE 1-24 #domain signal sequence #status predicted #label SIG\
25-111 #product cytochrome c6 #status predicted #label MAT\
28-101 #domain cytochrome c6 homology #label CYC\
38-41 #binding_site heme (Cys) (covalent) #status predicted\
42-82 #binding_site heme iron (His, Met) (axial ligands)
#status predicted
SUMMARY #length 111 #molecular-weight 11388 #checksum 5437

Query Match          2.6%; Score 6; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 2.80e+01;

```

```

Matches      6;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Db      41 CHLGR 46
      |||||
QY      221 CHLGR 226

RESULT      15
ENTRY      B25429      #type fragment
TITLE      T-cell receptor beta chain precursor V region (SUL73) - mouse
            (fragment)
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        02-Jun-1988 #sequence_revision 02-Jun-1988 #text_change
            30-May-1997
ACCESSIONS  B25429
REFERENCE    A94132
#authors     Behlke, M.A.; Chou, H.S.; Huppi, K.; Loh, D.Y.
#journal     Proc. Natl. Acad. Sci. U.S.A. (1986) 83:767-771
#title       Murine T-cell receptor mutants with deletions of beta-chain
            variable region genes.
#cross-references MUID:86121021
#accession   B25429
            #molecule_type mRNA
            #residues      1-120 #label BEH
CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology
KEYWORDS       T-cell receptor
FEATURE        11-120      #product T-cell receptor beta chain V region SUL73
            #status predicted #label MAT
SUMMARY        #length 120 #checksum 3746

Query Match      2.6%; Score 6; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.80e+01;
Matches      6;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Db      6 PCCGPG 11
      |||||
QY      27 PCCGPG 32

Search completed: Mon Jul 12 12:09:22 1999
Job time : 25 secs.

```

\*\*\*\*\*

W P S R L H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

on: Mon Jul 12 12:09:38 1999; MasPar time 8.03 Seconds  
803.090 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-911-423-2  
Description: (1-228) from US08911423.pep  
Perfect Score: 228  
Sequence: 1 MGAWMLYGVSMCLVLDLQ.....PEERGEOTEKCHLGGRRP 228

Scoring table:

TABLE unitprotable  
Gap 60  
Searched: 77977 seqs, 28268293 residues  
Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 3.389; Variance 0.374; scale 9.054

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	7	3.1	100	1	YD79_METJA	HYPOTHETICAL PROTEIN M	1.46e-01
2	7	3.1	218	1	Y522_METJA	HYPOTHETICAL PROTEIN M	1.46e-01
3	7	3.1	268	1	TRYP_STRGA	TRYPSIN-LIKE PROTEASE	1.46e-01
4	7	3.1	365	1	VNS2_AHSV9	NONSTRUCTURAL PROTEIN	1.46e-01
5	7	3.1	534	1	APG_ARATH	ANTER-SPECIFIC PROLINE	1.46e-01
6	7	3.1	575	1	SPD6_BACSU	STAGE VI SPOULATION P	1.46e-01
7	7	3.1	663	1	YD58_YEAST	HYPOTHETICAL 76.4 KD P	1.46e-01
8	7	3.1	773	1	MAK5_YEAST	ATP-DEPENDENT RNA HELI	1.46e-01
9	7	3.1	1858	1	P3K2_DICDI	PHOSPHATIDYLINOSITOL 3	1.46e-01
10	6	2.6	85	1	CYC6_ANANI	CYCLOCHROME C6 (SOLUBLE	1.48e-01
11	6	2.6	102	1	TRMA_SALTY	TRNA (URACIL-5-) METHY	1.48e-01
12	6	2.6	111	1	CYC6_SYNP7	CYCLOCHROME C6 PRECURSO	1.48e-01
13	6	2.6	119	1	B2MG_RAT	BETA-2-MICROGLOBULIN P	1.48e-01
14	6	2.6	159	1	HS21_SOYBN	17.9 KD CLASS II HEAT	1.48e-01
15	6	2.6	160	1	PHAA_MASIA	ALLOPHYCOCYANIN ALPHA	1.48e-01
16	6	2.6	160	1	PHAA_ANASP	ALLOPHYCOCYANIN ALPHA	1.48e-01
17	6	2.6	161	1	PHAC_FREDI	ALLOPHYCOCYANIN ALPHA	1.48e-01
18	6	2.6	164	1	YF18_FOWP1	HYPOTHETICAL 18.0 KD P	1.48e-01
19	6	2.6	182	1	YLH9_CAEEL	HYPOTHETICAL 21.0 KD P	1.48e-01
20	6	2.6	199	1	YECDC_ECOLI	HYPOTHETICAL 21.7 KD P	1.48e-01
21	6	2.6	204	1	ALAT7_ALTAL	MINOR ALLERGEN ALT A 7	1.48e-01
22	6	2.6	233	1	RS4E_HALMA	30S RIBOSOMAL PROTEIN	1.48e-01
23	6	2.6	239	1	YWIC_BACSU	HYPOTHETICAL 27.6 KD P	1.48e-01

24	1	HEX8_ADEG1	HEXON-ASSOCIATED PROTE	1.48e+01
25	1	PINK_ECOLI	PHOSPHONATES TRANSPORT	1.48e+01
26	1	TRYP_STRGR	TRYPSIN PRECURSOR (EC	1.48e+01
27	1	YR3_CAEEL	HYPOTHETICAL 30.8 KD P	1.48e+01
28	1	PSTA_METJA	PROBABLE PHOSPHATE TRA	1.48e+01
29	1	CYCG_RHOSH	DIHEME CYTOCHROME C-TY	1.48e+01
30	1	XYNB_STRLI	ENDO-1,4-BETA-XYLANASE	1.48e+01
31	1	BONZ_HUMAN	G PROTEIN-COUPLED RECE	1.48e+01
32	1	BONZ_HUMAN	G PROTEIN-COUPLED RECE	1.48e+01
33	1	ISLI_HUMAN	INSULIN GENE ENHANCER	1.48e+01
34	1	ISLI_BRARE	INSULIN GENE ENHANCER	1.48e+01
35	1	YJFR_ECOLI	HYPOTHETICAL 40.1 KD P	1.48e+01
36	1	YM97_YEAST	HYPOTHETICAL ZINC-TYPE	1.48e+01
37	1	YC25_YEAST	HYPOTHETICAL ZINC-TYPE	1.48e+01
38	1	VUL_HSV60	U1 PROTEIN.	1.48e+01
39	1	TRMA_ECOLI	TRNA (URACIL-5-) METHY	1.48e+01
40	1	CYB_ARTSF	CYCLOCHROME B (EC 1.10.	1.48e+01
41	1	YM81_YEAST	HYPOTHETICAL 44.9 KD P	1.48e+01
42	1	FTZ2_ARCFU	CELL DIVISION PROTEIN	1.48e+01
43	1	MLB_PSEPU	L(+)-MANDELATE DEHYDRO	1.48e+01
44	1	CYP7_YEAST	PEPTIDYL-PROLYL CIS-TR	1.48e+01
45	1	SYH2_SYNY3	PROBABLE HISTIDYL-TRNA	1.48e+01
46	1	ODPA_KLULA	PYRUVATE DEHYDROGENASE	1.48e+01
47	1	YGJU_HAEIN	HYPOTHETICAL SYMPORTER	1.48e+01
48	1	CAR4_CANAL	CANDIDAPEPSIN 4 PRECUR	1.48e+01
49	1	ODPA_YEAST	PYRUVATE DEHYDROGENASE	1.48e+01
50	1	PUR2_MYCLE	TRANSCRIPTIONAL ACTIVA	1.48e+01
51	1	GCN5_YEAST	PHOSPHORIBOSYLAMINE--G	1.48e+01
52	1	NU4M_CHLRE	NADH-UBIQUINONE OXIDOR	1.48e+01
53	1	GALP_ECOLI	GALACTOSE-PROTON SYMPO	1.48e+01
54	1	YHU6_YEAST	HYPOTHETICAL 51.1 KD P	1.48e+01
55	1	EXON_EBV	ALKALINE EXONUCLEASE (	1.48e+01
56	1	POL_AVIRE	POLYPROTEIN [CONTA	1.48e+01
57	1	YD92_METJA	CYTICHRONE P450 1A2 (E	1.48e+01
58	1	CP12_RAT	2-ISOPROPYLMALATE SYN	1.48e+01
59	1	LEU1_BACSU	HYPOTHETICAL BRP2 PRO	1.48e+01
60	1	POIB_ECOLI	PARAQUAT-INDUCIBLE PRO	1.48e+01
61	1	NRD1_YEAST	NRD1 PROTEIN.	1.48e+01
62	1	DPEP_SOLTU	4-ALPHA-GLUCANOTRANSF	1.48e+01
63	1	DPEP_SOLTU	HOMODECT DEFORMED PROT	1.48e+01
64	1	NU5M_LATCH	NADH-UBIQUINONE OXIDOR	1.48e+01
65	1	GP63_LEIGU	LEISHMANOLYSIN PRECURS	1.48e+01
66	1	KPCL_HUMAN	PROTEIN KINASE C, ETA	1.48e+01
67	1	KPCL_MOUSE	PROTEIN KINASE C, ETA	1.48e+01
68	1	KPCL_RAT	PROTEIN KINASE C, ETA	1.48e+01
69	1	YAS9_SCHPO	HYPOTHETICAL 82.4 KD P	1.48e+01
70	1	MBP1_KLULA	TRANSCRIPTION FACTOR M	1.48e+01
71	1	YRAJ_ECOLI	HYPOTHETICAL OUTER MEM	1.48e+01
72	1	P100_HSV60	LARGE STRUCTURAL PHOSP	1.48e+01
73	1	RB_HUMAN	RETINOBLASTOMA-ASSOCIA	1.48e+01
74	1	EAE_CITFR	INTIMIN (ATTACHING AND	1.48e+01
75	1	SACB_STRLI	LEVANSUCRASE PRECURSOR	1.48e+01
76	1	GUNZ_CLOS	ENDOGUCANASE Z PRECUR	1.48e+01
77	1	DPOL_VAVR	DNA POLYMERASE (EC 2.7	1.48e+01
78	1	VG07_BPT4	PHOSPHORIBOSYLAMINE--G	1.48e+01
79	1	GCSP_SOLTU	BASEPLATE STRUCTURAL P	1.48e+01
80	1	PGDR_HUMAN	GLYCINE DEHYDROGENASE	1.48e+01
81	1	EX5B_ECOLI	BETA PLATELET-DERIVED	1.48e+01
82	1	YQY3_CAEEL	EXODEOXYRIBONUCLEASE V	1.48e+01
83	1	SRF3_BACSU	HYPOTHETICAL 133.5 KD	1.48e+01
84	1	PUR2_DROME	SURFACTIN SYNTHETASE S	1.48e+01
85	1	PUR2_DROME	PHOSPHORIBOSYLAMINE--G	1.48e+01
86	1	NOS1_HUMAN	NITRIC-OXIDE SYNTHASE,	1.48e+01
87	1	CO3_LAMJA	COMPLEMENT C3 PRECURSO	1.48e+01
88	1	GLS1_YEAST	1,3-BETA-GLUCAN SYNTHA	1.48e+01
89	1	BRR2_YEAST	1,3-BETA-GLUCAN SYNTHA	1.48e+01
90	1	BRR2_YEAST	PRE-MRNA SPLICING HELI	1.48e+01
91	1	POLG_HPVA4	GENOME POLYPROTEIN [CO	1.48e+01
92	1	POLG_HPVA2	GENOME POLYPROTEIN [CO	1.48e+01
93	1	POLG_HPVA8	GENOME POLYPROTEIN [CO	1.48e+01
94	1	POLG_HPVA8	GENOME POLYPROTEIN [CO	1.48e+01
95	1	CIC5_HUMAN	N-TYPE CALCIUM CHANNEL	1.48e+01
96	1	CIC5_HUMAN	N-TYPE CALCIUM CHANNEL	1.48e+01

97 6 2.6 3010 1 POLG\_HCVJT GENOME POLYPROTEIN [CO 1.48e+01  
98 6 2.6 3011 1 POLG\_HCV1 GENOME POLYPROTEIN [CO 1.48e+01  
99 6 2.6 3660 1 DMD\_CHICK DYSTROPHIN  
100 6 2.6 3924 1 ANKE\_HUMAN ANKYRIN, BRAIN VARIANT 1.48e+01

## ALIGNMENTS

RESULT 1  
ID YD79\_METJA STANDARD; PRT; 100 AA.

AC Q58774;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE HYPOTHETICAL PROTEIN MJ1379.

GN MJ1379

OS METHANOCOCCUS JANNASCHII.

OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;

CC METHANOCOCCUS.

CC [1]

SEQUENCE FROM N.A.

RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE; 96337999.

RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii";

RL SCIENCE 273:1058-1073(1996).  
CC -!- SIMILARITY: BELONGS TO THE M.JANNASCHII MJ0126 / MJ0128 / MJ0141 /  
CC MJ0435 / MJ0504 / MJ1215 / MJ1217 / MJ1305 / MJ1379 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U67578; G1592331; -

DR TIGR; MJ1379; -

DE HYPOTHETICAL PROTEIN.

SEQUENCE 100 AA; 11819 MW; 7EC0FB79 CRC32;

Query Match 3.1%; Score 7; DB 1; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.46e-01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 35 RGEOTEE 41

QY 213 RGEOTEE 219

RESULT 2

ID Y522\_METJA STANDARD; PRT; 218 AA.  
AC Q57942;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE HYPOTHETICAL PROTEIN MJ0522.

GN MJ0522.

OS METHANOCOCCUS JANNASCHII.

OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;

CC METHANOCOCCUS.

CC [1]

SEQUENCE FROM N.A.

RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE; 96337999.  
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii";  
RL SCIENCE 273:1058-1073(1996).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U67502; G1510596; -

DR EMBL; U67502; G1591225; -

DR TIGR; MJ0522; -

KW HYPOTHETICAL PROTEIN: TRANSMEMBRANE.

FT TRANSMEM 19 39 POTENTIAL.

FT TRANSMEM 92 112 POTENTIAL.

FT TRANSMEM 124 144 POTENTIAL.

FT TRANSMEM 161 181 POTENTIAL.

FT TRANSMEM 196 216 POTENTIAL.

SQ SEQUENCE 218 AA; 23896 MW; 6022C576 CRC32;

Query Match 3.1%; Score 7; DB 1; Length 218;

Best Local Similarity 100.0%; Pred. No. 1.46e-01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 GDIVFGF 22

QY 92 GDIVFGF 98

RESULT 3

ID TRYP\_STRGA STANDARD; PRT; 268 AA.  
AC Q54179;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE TRYPSIN-LIKE PROTEASE PRECURSOR (EC 3.4.21.-).

OS STREPTOMYCES GLAUCEUS.

OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;

OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ETH 22794;

RA HINTERMANN G.;

RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U13770; G535440; -

DR PROSITE; PS00134; TRYPSIN\_HIS; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; 1.

DR PFAM; PF00089; trypsin; 1.

```
DR HSP: P00775; LSGT.
KW HYDROLASE; SERINE PROTEASE; ZYMOGEN; SIGNAL.
FT SIGNAL 1 41 POTENTIAL.
FT PROPEP 42 45 ACTIVATION PEPTIDE.
FT CHAIN 45 268 TRYPSIN.
FT ACT_SITE 82 82 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 127 127 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 67 83 BY SIMILARITY.
FT DISULFID 187 202 BY SIMILARITY.
FT DISULFID 213 242 BY SIMILARITY.
FT SITE 211 211 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 268 AA; 27550 MW; CFAD44FB CRC32;

Query Match 3.1%; Score 7; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.46e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 86 GSGNTR 92
|||||
37 GSGNTR 43

RESULT 4
ID VNS2_AHSV9 STANDARD; PRT; 365 AA.
AC P27279;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL PROTEIN NS2.
GN S8.
OS AFRICAN HORSE SICKNESS VIRUS 9 (AHSV-9) (AFRICAN HORSE SICKNESS VIRUS
OS (SEROTYPE 9)).
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ORBIVIRUS.
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE: 92024120.
RA VAN STADEN V., THERON J., GREYLING B.J., HUISMANS H., NEL L.H.; of
RT "A comparison of the nucleotide sequences of cognate NS2 genes of
RT three different orbiviruses.";
RL VIROLOGY 185:500-504(1991).
CC -!- FUNCTION: SSRNA-BINDING PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M69090; G210059; -.
CC PIR; A40788; MNXRAH.
CC NONSTRUCTURAL PROTEIN: RNA-BINDING.
SQ SEQUENCE 365 AA; 41193 MW; 8A3D3D11C CRC32;

Query Match 3.1%; Score 7; DB 1; Length 365;
Best Local Similarity 100.0%; Pred. No. 1.46e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 15 CVLIDLQ 21
|||||
14 CVLIDLQ 20

RESULT 5
ID APG-ARATH STANDARD; PRT; 534 AA.
AC P40602;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE ANTER-SPECIFIC PROLINE-RICH PROTEIN APG PRECURSOR.
GN APG.
```

```
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94004980.
RA ROBERTS M.R., FOSTER G.D., BLUNDELL R.P., ROBINSON S.W., KUMAR A.,
RA DRAPER J., SCOTT R.;
RT "Gametophytic and sporophytic expression of an anther-specific
RT Arabidopsis thaliana gene.";
RL PLANT J. 3:111-120(1993).
CC -!- TISSUE SPECIFICITY: FOUND IN SPOROPHYTIC AND GAMETOPHYTIC CELL
CC TYPES IN THE ANTER, ONLY IN MALE FERTILE PLANTS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING
CC MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING
CC MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
CC MATURATION.
CC -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X60377; G22599; -.
CC PIR; S21961; S21961.
CC PROSITE; PS01098; LIPASE_GDSL_SER; 1.
CC PFAM; PF00657; Lipase_GDSL; 1.
KW SIGNAL.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 534 ANTER-SPECIFIC PROLINE-RICH PROTEIN APG.
FT ACT_SITE 211 211 BY SIMILARITY.
FT ACT_SITE 511 511 POTENTIAL.
SQ SEQUENCE 534 AA; 57967 MW; 2D3EEF4A CRC32;

Query Match 3.1%; Score 7; DB 1; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.46e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 26 CIEFLT 32
|||||
164 CIEFLT 170

RESULT 6
ID SP6D_BACSU STANDARD; PRT; 575 AA.
AC P37963;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE STAGE VI SPORULATION PROTEIN D.
GN SPOVID.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93194796.
RA BEALL B., DRIKS A., LOSICK R., MORAN C.P. JR.;
RT "Cloning and characterization of a gene required for assembly of the
RT Bacillus subtilis spore coat.";
RL J. BACTERIOL. 175:1703-1716(1993).
CC -!- FUNCTION: REQUIRED FOR ASSEMBLY OF A NORMAL SPORE COAT. MAY BE A
CC COMPONENT OF THE INNERMOST LAYER OF THE SPORE COAT THAT AIDS IN
CC ITS ADHERENCE TO THE PRESPORE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION IS INITIATED AROUND THE SECOND
CC HOUR OF SPORULATION AND CONTINUES THROUGHOUT DEVELOPMENT. MAY BE
CC EXPRESSED PREDOMINANTLY IN THE MOTHER CELL.
CC -----
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; L07792; G143661; -;  
DR EMBL; Z99118; E1184060; -;  
DR SUBTILIST; BG10346; SPOVID.  
KW SPOULATATION.  
SQ SEQUENCE 575 AA; 64976 MW; 5E99BC50 CRC32;

Query Match 3.1%; Score 7; DB 1; Length 575;  
Best Local Similarity 100.0%; Pred. No. 1.46e-01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DR 168 SVVEEPG 174  
|||||||  
22 SVVEEPG 28

RESULT 7  
ID YD58 YEAST STANDARD; PRT; 663 AA.  
AC Q03281;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE HYPOHETICAL 76.4 K D PROTEIN IN GUK1-MFAL INTERGENIC REGION.  
GN YD458C OR D8035.2.  
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACHAROMYCETALES;  
OC SACHAROMYCETACEAE; SACHAROMYCES.  
RN [1]

RA DIETRICH F.S., MULLIGAN J., ALLEN E., ARAUJO R., AVILES E.,  
RA BERNO A., CARPENTER J., CHEN E., CHERRY J.M., CHUNG E., DUNCAN M.,  
RA HUNICKE-SMITH S., HYMAN R., KOMP C., LASHKARI D., LEW H., LIN D.,  
RA MOSEDALE D., NAKAHARA K., NAMATH A., OEFNER P., OH C., PETEL F.X.,  
RA ROBERTS D., SCHRAMM S., SCHROEDER M., SHOGREN T., SHROFF N.,  
RA WINANT A., YELTON M., BOTSTEIN D., DAVIS R.W.;  
RL SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- SIMILARITY: TO YEAST YML034W (IN THE N-TERMINAL) AND YML033W  
(IN THE C-TERMINAL).  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U33050; G927762; -;  
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
FT DOMAIN 85 90 POLY-SER.  
FT TRANSMEM 318 338 POTENTIAL.  
SQ SEQUENCE 663 AA; 76376 MW; 4DCA2D86 CRC32;

Query Match 3.1%; Score 7; DB 1; Length 663;  
Best Local Similarity 100.0%; Pred. No. 1.46e-01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DR 605 FLTTVOL 611  
|||||||  
QY 167 FLTTVOL 173

RESULT 8  
ID MAK5 YEAST STANDARD; PRT; 773 AA.  
AC P38112;  
DT 01-OCT-1994 (REL. 30, CREATED)

DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE ATP-DEPENDENT RNA HELICASE MAK5.  
GN MAK5 OR YBR142W OR YBR1119.  
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACHAROMYCETALES;  
OC SACHAROMYCETACEAE; SACHAROMYCES.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE; 95274325.  
RA ZAGULSKI M., BECAM A.-M., GRZYBOWSKA E., LACROUTE F., MIGDALSKI A.,  
RA SLONIMSKI P.P., SOKOLOWSKA B., HERBERT C.J.;  
RT "The sequence of 12.5 kb from the right arm of chromosome II predicts  
RT a new N-terminal sequence for the IRA1 protein and reveals two new  
RT genes, one of which is a DEAD-box helicase."  
RL YEAST 10:1227-1234(1994).  
RN [2]

RP SEQUENCE OF 770-773 FROM N.A.  
RC STRAIN-S288C;  
RA ENTIAN K.-D., KOETTER P., ROSE M., BECKER J., GREY M., LI Z.,  
RA NIEGEMANN E., SCHENK-GROENINGER R., SERVOS J., WEHNER E.,  
RA WOLTER R., BRENDL M., BAUER J., BRAUN H., DERN K., DUESTERHUS S.,  
RA GRUENBEIN R., HEDGES D., KIESAU P., KOROL S., KREMS B., PROFT M.,  
RA SIEGERS K., BAUR A., BOLES E., MIOSGA T.,  
RA SCHAAFF-GERSTENSCHLAGER I., ZIMMERMANN F.K.;  
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: INVOLVED IN MAINTENANCE OF DSRNA KILLER PLASMID.  
CC -!- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; Z36011; G536442; -;  
DR EMBL; X78937; G547578; -;  
DR PIR; S46011; S46011.  
DR SGD; L0000579; MAK5.  
DR PROSITE; PS00039; DEAD\_ATP\_HELICASE; 1.  
DR PFAM; PF00270; DEAD; 1.  
DR PFAM; PF00271; Helicase\_C; 1.  
DR HELICASE; ATP-BINDING.  
FT NP\_BIND 215 222 ATP (POTENTIAL).  
FT SITE 333 336 DEAD BOX.  
SQ SEQUENCE 773 AA; 87048 MW; 88D14EDA CRC32;

Query Match 3.1%; Score 7; DB 1; Length 773;  
Best Local Similarity 100.0%; Pred. No. 1.46e-01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DR 458 FLTMFPG 464  
|||||||  
QY 127 FLTMFPG 133

RESULT 9  
ID P3K2\_DICDI STANDARD; PRT; 1858 AA.  
AC P54674;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE PHOSPHATIDYLINOSITOL 3-KINASE 2 (EC 2.7.1.137) (PI3-KINASE)  
DE (PTDINS-3-KINASE) (PI3K).  
GN PIK2 OR PIK2.  
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).  
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AX3;

RX MEDLINE: 96009592.  
RA ZHOU K., TAKEGAWA K., EMR S.D., FIRTEL R.A.;  
RT "A phosphatidylinositol (PI) kinase gene family in Dictyostelium  
RT discoidium: biological roles of putative mammalian p110 and yeast  
RT vps34p PI 3-kinase homologs during growth and development.";  
RL MOL. CELL. BIOL. 15:5645-5656(1995).  
CC -!- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL = ADP +  
CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE.  
CC -!- SIMILARITY: BELONGS TO THE P13/PI4-KINASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U23477; G73322; .  
DI DICTYDB: DD01100; PIK8.  
KW PROSITE: PS00915; PI3\_4\_KINASE\_1; 1.  
FT PROSITE: PS00916; PI3\_4\_KINASE\_2; 1.  
DR PFAM: PF00454; PI3\_P14\_Kinase; 1.  
DR PFAM: PF00613; PI3Ka; 1.  
DR PFAM: PF00792; PI3K\_C2; 1.  
DR PFAM: PF00794; PI3K\_rbd; 1.  
KW TRANSFERASE; KINASE; MULTIGENE FAMILY.  
FT DOMAIN 34 40 POLY-GLY.  
FT DOMAIN 166 172 POLY-SER.  
FT DOMAIN 185 226 POLY-ASN.  
FT DOMAIN 227 235 POLY-THR.  
FT DOMAIN 246 253 POLY-SER.  
FT DOMAIN 259 268 POLY-ASN.  
FT DOMAIN 294 303 POLY-THR.  
FT DOMAIN 361 364 POLY-SER.  
FT DOMAIN 369 384 POLY-SER.  
FT DOMAIN 425 429 POLY-GLN.  
FT DOMAIN 439 444 POLY-THR.  
FT DOMAIN 445 454 POLY-SER.  
FT DOMAIN 562 570 POLY-GLY.  
FT DOMAIN 715 727 POLY-THR.  
FT DOMAIN 982 990 POLY-GLN.  
FT DOMAIN 1015 1049 POLY-ASN.  
FT DOMAIN 1598 1858 PI3K/PI4K.  
SQ SEQUENCE 1858 AA; 203945 MW; 228971F5 CRC32;  
  
Query Match 3.1%; Score 7; DB 1; Length 1858;  
Best Local Similarity 100.0%; Pred. No. 1.46e-01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
DB 1351 VOLSAD 1357  
|||||||  
QY 196 VOLSAD 202  
  
RESULT 10  
ID CYC\_ANNANI STANDARD; PRT: 85 AA.  
AC P07497;  
DT 01-APR-1988 (REL. 07, CREATED)  
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE CYTOCHROME C6 (SOLUBLE CYTOCHROME F) (CYTOCHROME C553).  
GN PETJ.  
OS ANACYSTIS NIDULANS.  
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCOCCUS.  
RN [1]  
RP SEQUENCE.  
RA MARGOLIASH E.;  
RL UNPUBLISHED RESULTS, CITED BY:  
RL DICKERSON R.E.;  
RL (IN) SIGMAN D.S., BRAZIER M.A.B. (EDS.);  
RL THE EVOLUTION OF PROTEIN STRUCTURE AND FUNCTION, PP.173-202, ACADEMIC  
RL PRESS, NEW YORK AND LONDON (1980).

CC -!- FUNCTION: CYTOCHROME C6 IS A MONOHEME MONOMER. IT FUNCTIONS AS AN  
CC ELECTRON CARRIER BETWEEN MEMBRANE-BOUND CYTOCHROME F AND P700 IN  
CC THE PHOTOPHOSPHORYLATION CHAIN IN CHLOROPLASTS AND ALGAE. IT  
CC SUBSTITUTES FOR PLASTOCYANIN IN COPPER-DEFICIENT BLUE-GREEN ALGAE  
CC AND IN THE CHLOROPLASTS OF SOME EUKARYOTE ALGAE.  
DR PIR; A05180; A05180.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
DR PFAM: PF00034; cytochrome\_c; 1.  
DR HSP: P08197; 1CYL.  
KW ELECTRON TRANSPORT; PHOTOSYNTHESIS; HEME.  
FT BINDING 14 14 HEME (COVALENT).  
FT BINDING 17 17 HEME (COVALENT).  
FT METAL 18 18 IRON (HEME AXIAL LIGAND).  
FT METAL 58 58 IRON (HEME AXIAL LIGAND).  
SQ SEQUENCE 85 AA; 8743 MW; 80D6CE2D CRC32;  
  
Query Match 2.6%; Score 6; DB 1; Length 85;  
Best Local Similarity 100.0%; Pred. No. 1.48e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
DB 17 CHLGR 22  
|||||  
QY 221 CHLGR 226  
  
RESULT 11  
ID TRMA\_SALTY STANDARD; PRT: 102 AA.  
AC P22038;  
DT 01-AUG-1991 (REL. 19, CREATED)  
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE TRNA (URACIL-5-)-METHYLTRANSFERASE (EC 2.1.1.35) (TRNA(M-5-U54))-  
DE METHYLTRANSFERASE) (RUMT) (FRAGMENT).  
GN TRMA.  
OS SALMONELLA TYPHIMURIUM.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
OC SALMONELLA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 91154132.  
RA GUSTAFSSON C., LINDSTROM P.H.R., HAGERVALL T.G., ESBERG K.B.,  
RA BJORK G.R.;  
RT "The trna promoter has regulatory features and sequence elements in  
RT common with the trna P1 promoter family of Escherichia coli.";  
RT J. BACTERIOL. 173:1757-1764(1991).  
CC -!- FUNCTION: CATALYZES THE FORMATION OF 5-METHYL-URIDINE AT POSITION  
CC 54 (M-5-U54) IN ALL TRNA.  
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA =  
CC S-ADENOSYL-L-HOMOCYSTEINE + TRNA CONTAINING THYMINE.  
CC -!- INDUCTION: GROWTH RATE-DEPENDENT REGULATION OF TRANSCRIPTION. IS A  
CC NOVEL EXAMPLE OF A MRNA REGULATED THROUGH A MECHANISM SIMILAR TO  
CC THAT OF A STABLE RNA (RRNA).  
CC -!- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRMA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M57569; G552024; .  
DR STYGENE; SG10394; TRNA.  
DR PROSITE; PS01230; TRNA\_1; PARTIAL.  
DR PROSITE; PS01231; TRNA\_2; PARTIAL.  
KW TRANSFERASE; METHYLTRANSFERASE; TRNA PROCESSING.  
FT NON\_TER 102 102  
SQ SEQUENCE 102 AA; 11853 MW; 39451F6E CRC32;  
  
Query Match 2.6%; Score 6; DB 1; Length 102;  
Best Local Similarity 100.0%; Pred. No. 1.48e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 6 LPTEQY 11
    |||||
QY 146 LPTEQY 151

RESULT 12
ID CYC6_SYP7 STANDARD; PRT; 111 AA.
AC P25935;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE CYTOCHROME C6 PRECURSOR (SOLUBLE CYTOCHROME F) (CYTOCHROME C553).
GN PETJ OR CVTA.
OS SYNECHOCOCCUS SP. (STRAIN PCC 7942) (ANACYSTIS NIDULANS R2).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE; 93005680.
LAUDENBACH D.E., HERBERT S.K., MCDOWELL C., FORK D.C., GROSSMAN A.R.,
STRAUS N.A.;
"Cytochrome c-553 is not required for photosynthetic activity in the
cyanobacterium Synechococcus.";
PLANT CELL 2:913-924(1990).
-!- FUNCTION: CYTOCHROME C6 IS A MONOHEME MONOMER. IT FUNCTIONS AS AN
ELECTRON CARRIER BETWEEN MEMBRANE-BOUND CYTOCHROME F AND P700 IN
THE PHOTOPHOSPHORYLATION CHAIN IN CHLOROPLASTS AND ALGAE. IT
SUBSTITUTES FOR PLASTOCYANIN IN COPPER-DEFICIENT BLUE-GREEN ALGAE
AND IN THE CHLOROPLASTS OF SOME EUKARYOTE ALGAE.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; S44426; G256652; -.
PIR; JQ1083; JQ1083.
PROSITE; PS00190; CYTOCHROME_C; 1.
PFAM; PF00034; cytochrome_c; 1.
HSP; P56534; 1C6S.
ELECTRON TRANSPORT; PHOTOSYNTHESIS; HEME; SIGNAL.
FT SIGNAL 1 24 BY SIMILARITY.
FT CHAIN 25 111 CYTOCHROME C6.
FT BINDING 38 38 HEME (COVALENT) (BY SIMILARITY).
FT BINDING 41 41 HEME (COVALENT) (BY SIMILARITY).
FT METAL 42 42 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT METAL 82 82 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 111 AA; 11388 MW; 803579DC CRC32;

Query Match 2.6%; Score 6; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.48e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 41 CHLGR 46
    |||||
QY 221 CHLGR 226

RESULT 13
ID B2MG_RAT STANDARD; PRT; 119 AA.
AC P07151;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE BETA-2-MICROGLOBULIN PRECURSOR.
GN B2M.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ROSENTHIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR;
RX MEDLINE; 88015621.
RA MAUXION F., KRESS M.;
RT "Nucleotide sequence of rat beta 2-microglobulin cDNA.";
RL NUCLEIC ACIDS RES. 15:7638-7638(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-RTIO STOCK; TISSUE=SPLENOCYTE;
RA LE ROLLE A.F., BUTCHER G.W., JOLY E.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 21-39.
RC STRAIN-WISTAR;
RX MEDLINE; 89097257.
RA SIMISTER N.E., MOSTOV K.E.;
RT "An Fc receptor structurally related to MHC class I antigens.";
RL NATURE 337:184-187(1989).
RN [4]
RP SEQUENCE OF 65-119 FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;
RA COLE T., DICKSON P.W., ESNAUD F., AVERILL F., RISBRIDGER G.,
GAUTHIER F., SCHREIBER G.;
RT "The cDNA structure and expression analysis of the genes for the
cysteine proteinase inhibitor cystatin C and for beta 2-microglobulin
in rat brain.";
RL EUR. J. BIOCHEM. 186:35-42(1989).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE; 95059482.
RA BURMEISTER W.P., HUBER A.H., BJORKMAN P.J.;
RT "Crystal structure of the complex of rat neonatal Fc receptor with
Fc.";
RL NATURE 372:379-383(1994).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE; 98154319.
RA VAUGHN D.E., BJORKMAN P.J.;
RT "Structural basis of pH-dependent antibody binding by the neonatal Fc
receptor.";
RL STRUCTURE 6:63-73(1998).
CC -!- FUNCTION: BETA-2-MICROGLOBULIN IS THE BETA-CHAIN OF MAJOR
HISTOCOMPATIBILITY COMPLEX CLASS I MOLECULES.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Y00441; G55808; -.
EMBL; Y08531; E274136; -.
EMBL; X16956; G818019; -.
PIR; A26842; A26842.
PIR; S10087; S10087.
PDB; 1FRT; 14-FEB-95.
PDB; 3FRU; 10-JUN-98.
PDB; 3FRT; 10-JUN-98.
PFAM; PF00047; Ig; 1.
MHC I; SIGNAL; 3D-STRUCTURE.
FT SIGNAL 1 20
FT CHAIN 21 119 BETA-2-MICROGLOBULIN.
FT DISULFID 45 100
SQ SEQUENCE 119 AA; 13720 MW; 2C30E8C3 CRC32;

Query Match 2.6%; Score 6; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 1.48e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Db 6 TVIFLV 11  
 QY 155 TVIFLV 160

RESULT 14  
 ID HS21\_SOYBN STANDARD; PRT; 159 AA.  
 AC P05477;  
 DT 01-NOV-1988 (REL. 09, CREATED)  
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
 DE 17.9 KD CLASS II HEAT SHOCK PROTEIN.  
 GN HSP17.9-D.

OS GLYCINE MAX (SOYBEAN).  
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
 OC FABALES; FABACEAE; PAPILIONOIDEAE; GLYCINE.

RN [1]  
 RP SEQUENCE FROM N.A.  
 STRAIN=CV, CORSOY;  
 MEDLINE; 88172502.  
 RASCHKE E., BAUMANN G., SCHOEFL F.;  
 RT "Nucleotide sequence analysis of soybean small heat shock protein  
 genes belonging to two different multigene families.";  
 RL J. MOL. BIOL. 199;549-557(1988).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)  
 CC FAMILY.

CC -!- SIMILARITY: PLANTS HAVE FOUR CLASSES OF HSP20: CLASS I AND II  
 CC WHICH ARE CYTOPLASMIC; CLASS III WHICH IS IN THE CHLOROPLAST  
 CC AND CLASS IV WHICH IS IN THE ENOMEMBRANE. THIS PROTEIN BELONGS  
 CC TO CLASS II.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X071159; G18652;  
 DR PIR; S01859; S01859;  
 DR PROSITE; PS01031; HSP20; 1.  
 DR PFAM; PF00011; HSP20; 1.  
 KW HEAT SHOCK; MULTIGENE FAMILY.  
 SO SEQUENCE 159 AA; 17878 MW; 9837ADE2 CRC32;

Query Match 2.68; Score 6; DB 1; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 1.48e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 30 NKTHNA 35  
 QY 134 NKTHNA 139

RESULT 15  
 ID PHAA\_MASLA STANDARD; PRT; 160 AA.  
 AC P00315;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE ALLOPHYCOCYANIN ALPHA CHAIN.  
 GN APCA.

OS MASTIGOCLADUS LAMINOSUS (FISCHERELLA SP.).  
 OC BACTERIA; CYANOBACTERIA; STIGONEMATALES; FISCHERELLA.

RN [1]  
 RP SEQUENCE.

RX MEDLINE; 82005802.

RA SIDLER W., GYSI J., ISKER E., ZUBER H.;

RT "The complete amino acid sequence of both subunits of  
 RT allophycocyanin, a light harvesting protein-pigment complex from the

RT cyanobacterium Mastigocladus laminosus.";  
 RL HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 362:611-628(1981).  
 CC -!- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN  
 CC FROM THE PHYCOBILIPROTEIN COMPLEX. ALLOPHYCOCYANIN HAS A MAXIMUM  
 CC ABSORPTION AT APPROXIMATELY 650 NANOMETERS.  
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
 CC PIR; A00325; AFMWA.  
 DR PFAM; PF00502; Phycobilisome; 1.  
 DR HSP; P07119; ICPC.  
 KW PHYCOBILISOME; ELECTRON TRANSPORT; PHOTOSYNTHESIS; BILE PIGMENT.  
 FT BINDING 80 80 PHYCOCYANOBILIN CHROMOPHORE.  
 SO SEQUENCE 160 AA; 17121 MW; 9E4214E3 CRC32;

Query Match 2.68; Score 6; DB 1; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 1.48e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 140 LSAEDA 145  
 QY 198 LSAEDA 203

Search completed: Mon Jul 12 12:09:57 1999  
 Job time : 19 secs.

\*\*\*\*\*  
WIRE (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
on: Mon Jul 12 12:13:01 1999; MasPar time 12.25 Seconds  
Tabular output not generated. 788.318 Million cell updates/sec

Title: >US-08-911-423-4  
Description: (1-241) from US08911423.pep  
Perfect Score: 241  
Sequence: 1 MAQHGAMGAFRALCGLALLC.....EEERGERSAEKRGDLWV 241

Scoring table: TABLE unitprotable  
Gap 60  
Searched: 122810 seqs, 40068593 residues  
Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 3.477; Variance 0.464; scale 7.495

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Sult No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	8	3.3	241	2	E70470	conserved hypotheticala	3.57e-02
2	8	3.3	411	2	S61245	probable virion glyco	3.57e-02
3	8	3.3	509	2	H70597	probable membrane pro	3.57e-02
4	8	3.3	782	1	FAHUP	gelsolin precursor, p	3.57e-02
5	7	2.9	28	2	I48349	fibronectin - mouse (	2.50e+00
6	7	2.9	120	2	B25429	T-cell receptor beta	2.50e+00
7	7	2.9	133	2	A71173	hypothetical protein	2.50e+00
8	7	2.9	144	1	UTCAB	gonadotropin beta cha	2.50e+00
9	7	2.9	206	2	E69961	N-acetylmuramyl-L-al	2.50e+00
10	7	2.9	232	2	S58353	Cd1b protein - sheep	2.50e+00
11	7	2.9	256	2	D69845	thiamin biosynthesis	2.50e+00
12	7	2.9	268	2	B42424	chitinase (EC 3.2.1.1	2.50e+00
13	7	2.9	274	2	B45887	leukocyte differentia	2.50e+00
14	7	2.9	286	2	S07532	puff II/9-1 protein p	2.50e+00
15	7	2.9	286	2	S07533	puff II/9A-2 protein	2.50e+00
16	7	2.9	297	2	I38517	syntaxin - human	2.50e+00
17	7	2.9	297	2	S52726	syntaxin-4 - human	2.50e+00
18	7	2.9	324	2	S51001	transforming protein	2.50e+00
19	7	2.9	332	2	S63660	NADH dehydrogenase (u	2.50e+00
20	7	2.9	333	2	S47246	gene Cdl protein - sh	2.50e+00
21	7	2.9	335	1	HLHUR3	T-cell surface glycop	2.50e+00
22	7	2.9	350	2	T03485	ornithine cyclodeamin	2.50e+00
23	7	2.9	352	2	S76078	hypothetical protein	2.50e+00

24	7	2.9	359	2	A56549	cell-cell signaling m	2.50e+00
25	7	2.9	377	2	G71341	conserved hypotheticala	2.50e+00
26	7	2.9	401	2	A36961	pilin biogenesis prot	2.50e+00
27	7	2.9	447	2	S37048	cysteine proteinase -	2.50e+00
28	7	2.9	474	2	D89400	pantothenate permease	2.50e+00
29	7	2.9	508	1	GVGNFR	spike glycoprotein G	2.50e+00
30	7	2.9	524	1	VGVNVC	spike glycoprotein G	2.50e+00
31	7	2.9	598	2	D71391	NADH dehydrogenase (u	2.50e+00
32	7	2.9	603	1	DNHUNS	NADH dehydrogenase (u	2.50e+00
33	7	2.9	641	2	S63645	NADH dehydrogenase (u	2.50e+00
34	7	2.9	653	2	S23667	cc11 protein - Rhodob	2.50e+00
35	7	2.9	662	2	S62707	NADH dehydrogenase (u	2.50e+00
36	7	2.9	666	2	S59109	NADH dehydrogenase (u	2.50e+00
37	7	2.9	675	2	S53831	NADH dehydrogenase (u	2.50e+00
38	7	2.9	685	2	S68303	sel-1 protein - Caeno	2.50e+00
39	7	2.9	713	2	B38992	caderin 13 precursor	2.50e+00
40	7	2.9	750	2	S73829	probable ABC transpor	2.50e+00
41	7	2.9	830	2	S08100	probable membrane pro	2.50e+00
42	7	2.9	1216	2	JW0105	synaptotagmin 2 alpha	2.50e+00
43	7	2.9	1220	2	A56136	jagged protein precu	2.50e+00
44	7	2.9	1469	2	A55095	chromosome condensati	2.50e+00
45	7	2.9	2477	2	S14428	fibronectin precursor	2.50e+00
46	6	2.5	25	2	A60741	insulin-like growth f	1.09e+02
47	6	2.5	63	2	P50135	H-2 class I histocomp	1.09e+02
48	6	2.5	63	2	P50136	H-2 class I histocomp	1.09e+02
49	6	2.5	90	2	JN0492	hypothetical 9.7K pro	1.09e+02
50	6	2.5	105	2	S54732	uvil5 protein - fissi	1.09e+02
51	6	2.5	146	1	HBTX1	hemoglobin beta-1 cha	1.09e+02
52	6	2.5	186	2	I61783	sodium channel beta 2	1.09e+02
53	6	2.5	215	2	A57843	sodium channel beta 2	1.09e+02
54	6	2.5	215	2	J1340	hypothetical protein	1.09e+02
55	6	2.5	223	2	EC5266	sigma receptor 1 - hu	1.09e+02
56	6	2.5	223	3	C40049	virC-region protein y	1.09e+02
57	6	2.5	226	2	JN0464	insulin-like growth f	1.09e+02
58	6	2.5	238	2	I48605	insulin-like growth f	1.09e+02
59	6	2.5	250	2	E71402	hypothetical protein	1.09e+02
60	6	2.5	270	2	D71301	hypothetical protein	1.09e+02
61	6	2.5	281	2	H71308	hypothetical protein	1.09e+02
62	6	2.5	321	2	S55004	NADH dehydrogenase (u	1.09e+02
63	6	2.5	342	2	S76463	hypothetical protein	1.09e+02
64	6	2.5	350	2	I55214	salivary proline-rich	1.09e+02
65	6	2.5	353	2	B40891	GRP-binding protein G	1.09e+02
66	6	2.5	354	2	S33309	GRP-binding regulator	1.09e+02
67	6	2.5	359	1	RGHUGY	probable transposase	1.09e+02
68	6	2.5	371	2	S04880	3beta-hydroxy-Delta5-	1.09e+02
69	6	2.5	373	1	DEBOHS	coat protein VP1 - mo	1.09e+02
70	6	2.5	384	1	VVP13	chalcone synthase hom	1.09e+02
71	6	2.5	385	2	T01332	probable transcriptio	1.09e+02
72	6	2.5	389	2	G70810	hypothetical protein	1.09e+02
73	6	2.5	408	2	B71272	citrate carrier prote	1.09e+02
74	6	2.5	434	2	JQ0576	probable B-alanine gl	1.09e+02
75	6	2.5	451	2	D71327	hypothetical protein	1.09e+02
76	6	2.5	471	2	S05392	hypothetical protein	1.09e+02
77	6	2.5	500	2	T03387	polyamine oxidase (EC	1.09e+02
78	6	2.5	507	3	JG0165	LAR1 protein - Human	1.09e+02
79	6	2.5	510	2	G71365	probable carnitine tr	1.09e+02
80	6	2.5	526	2	G71126	probable alanine tran	1.09e+02
81	6	2.5	528	2	PC4025	intercellular adhesio	1.09e+02
82	6	2.5	538	2	S54015	C-22 sterol desaturas	1.09e+02
83	6	2.5	540	2	S21825	vicillin-like storage	1.09e+02
84	6	2.5	593	2	S49525	glycoprotein G - sm1	1.09e+02
85	6	2.5	683	1	A23690	protein kinase (EC 2.	1.09e+02
86	6	2.5	696	2	A29635	transcription factor	1.09e+02
87	6	2.5	729	2	A30904	anthranilate synthase	1.09e+02
88	6	2.5	785	2	S63652	hypothetical protein	1.09e+02
89	6	2.5	788	2	S07477	regulatory protein Sp	1.09e+02
90	6	2.5	788	2	F71908	ribonucleoside-diphos	1.09e+02
91	6	2.5	788	2	S25287	Spl protein - rat	1.09e+02
92	6	2.5	807	2	S40458	ribosomal protein S3	1.09e+02
93	6	2.5	1173	2	I50620	prockr2 - chicken (fr	1.09e+02
94	6	2.5	1840	1	CHRTM1	sodium channel protei	1.09e+02
95	6	2.5	1872	2	T00339	hypothetical protein	1.09e+02
96	6	2.5	1938	2	I37453	host cell factor C1 -	1.09e+02

97 6 2.5 1944 2 A51117 tsq24 protein - mouse 1.09e+02  
98 6 2.5 2035 2 A56088 host cell factor C1 p 1.09e+02  
99 6 2.5 2035 2 A40718 host cell factor C1 p 1.09e+02  
100 6 2.5 3020 2 A43932 mucin 2 precursor, in 1.09e+02

## ALIGNMENTS

RESULT 1  
ENTRY conserved hypothetical protein aq\_1986 - Aquifex aeolicus  
TITLE conserved hypothetical protein aq\_1986 - Aquifex aeolicus  
ORGANISM #formal\_name Aquifex aeolicus  
DATE 08-May-1998 #sequence\_revision 08-May-1998 #text\_change  
08-May-1998  
ACCESSIONS E70470 #type complete  
REFERENCE A70300 conserved hypothetical protein aq\_1986 - Aquifex aeolicus  
#authors Lennox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;  
Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,  
J.M.; Olson, G.J.; Swanson, R.V.  
#journal Nature (1998) 392:353-358  
#title The complete genome of the hyperthermophilic bacterium  
Aquifex aeolicus.  
#cross-references MUID:98196666  
#accession E70470  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
##molecule\_type DNA  
##residues 1-241 #label AOF  
##cross-references GB:AE000766; NID:g2984216; PID:g2984233; GB:AE000657  
##experimental\_source strain VF5  
GENETICS  
#gene aq\_1986  
SUMMARY #length 241 #molecular-weight 27514 #checksum 451

Query Match 3.3%; Score 8; DB 2; Length 241;  
Best Local Similarity 100.0%; Pred. No. 3.57e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 38 ALLCALS 45  
QY 17 ALLCALS 24

RESULT 2  
ENTRY #type complete  
TITLE probable varion glycoprotein M (gm) - bovine herpesvirus 1  
ORGANISM #formal\_name bovine herpesvirus 1  
DATE 18-Sep-1997 #sequence\_revision 18-Sep-1997 #text\_change  
21-Aug-1998  
ACCESSIONS S61245  
REFERENCE S61233  
#authors Vileck, C.; Benes, V.; Lu, Z.; Kutish, G.F.; Paces, V.; Rock,  
D.; Lettoworth, G.J.; Schwytzer, M.  
#submission Submitted to the EMBL Data Library, January 1995  
#description Nucleotide sequence analysis of a 30-Kb region of the bovine  
herpesvirus 1 genome which exhibits a colinear gene  
arrangement with the UL21 to UL4 genes of herpes simplex  
virus.  
#accession S61245  
#status preliminary  
##molecule\_type DNA  
##residues 1-411 #label VLC  
##cross-references EMBL:248053; NID:g971311; PID:g971324  
CLASSIFICATION #superfamily herpesvirus 51K protein  
SUMMARY #length 411 #molecular-weight 43029 #checksum 571

Query Match 3.3%; Score 8; DB 2; Length 411;  
Best Local Similarity 100.0%; Pred. No. 3.57e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 93 VVLLAVAA 100  
QY 17 VVLLAVAA 100

QY 168 VVLLAVAA 175  
RESULT 3  
ENTRY #type complete  
TITLE probable membrane protein - Mycobacterium tuberculosis  
(strain H37RV)  
ORGANISM #formal\_name Mycobacterium tuberculosis  
DATE 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change  
17-Jul-1998  
ACCESSIONS H70597  
REFERENCE A70500  
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,  
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry,  
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;  
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;  
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;  
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; R.,  
Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;  
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;  
Taylor, K.; Whitehead, S.; Barrell, B.G.  
#journal Nature (1998) 393:537-544  
#title Deciphering the biology of Mycobacterium tuberculosis from  
the complete genome sequence.  
#cross-references MUID:98295987  
#accession H70597  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
##molecule\_type DNA  
##residues 1-509 #label COL  
##cross-references GB:294121; GB:AL123456; NID:g3261736; PID:e312290;  
PID:g1944601  
##experimental\_source strain H37RV  
GENETICS  
#gene RV3887C  
SUMMARY #length 509 #molecular-weight 53278 #checksum 6762

Query Match 3.3%; Score 8; DB 2; Length 509;  
Best Local Similarity 100.0%; Pred. No. 3.57e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 205 VVLLAVAA 212  
QY 168 VVLLAVAA 175

RESULT 4  
ENTRY #type complete  
TITLE FAHUP  
ALTERNATE\_NAMES gelolin precursor, plasma - human  
CONTAINS actin-depolymerizing factor; brevini  
ORGANISM amyloid protein, I2K; gelolin, cytosolic  
#formal\_name Homo sapiens #common\_name man  
DATE 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change  
26-Feb-1999  
ACCESSIONS A03011; A38797; S03073; A60849; A36029; A61263; A34562;  
A34137  
REFERENCE A93383  
#authors Kwiatkowski, D.J.; Stosel, T.P.; Orkin, S.H.; Mole, J.E.;  
Colten, H.R.; Yin, H.L.  
#journal Nature (1986) 323:455-458  
#title Plasma and cytoplasmic gelolins are encoded by a single gene  
and contain a duplicated actin-binding domain.  
#cross-references MUID:87014807  
#accession A03011  
##molecule\_type mRNA  
##residues 1-782 #label KWI  
##cross-references EMBL:X04412; NID:g35447; PID:g736249  
#accession A38797  
##molecule\_type protein  
##residues 28-52; 178-194; 'XX', 197-279-290; 'S', 292-294; 'XX', 297-303;  
434-449; 'XX', 452-453; 'X', 455-459; 538-542; 'XX', 545-581  
#label KW2

```
REFERENCE
#authors Kwiatkowski, D.J.; Mehl, R.; Yin, H.L.
#journal J. Cell Biol. (1998) 106:375-384
#title Genomic organization and biosynthesis of secreted and cytoplasmic forms of gelsolin.
#cross-references MUID:88115587
#accession S03073
#status translation not shown
#molecule_type DNA
#residues 1-116 #label KW3
#cross-references EMBL:X07065
#note I-Met is the initiator for plasma gelsolin. 52-Met is the initiator for cytosolic gelsolin

REFERENCE
#authors Bryan, J.; Hwo, S.
#journal J. Cell Biol. (1986) 102:1439-1446
#title Definition of an N-terminal actin-binding domain and a C-terminal Ca(2+) regulatory domain in human brevin.
#cross-references MUID:86168493
#accession A60849
#molecule_type protein
#residues 'X',52-55,'X',57-59,'X',61-62 #label BRY
#authors Vandekerckhove, J.; Bauw, G.; Vancompernelle, K.; Honore, B.; Celis, J.
#journal J. Cell Biol. (1990) 111:95-102
#title Comparative two-dimensional gel analysis and microsequencing identifies gelsolin as one of the most prominent downregulated markers of transformed human fibroblast and epithelial cells.
#cross-references MUID:90307803
#accession A36029
#molecule_type protein
#residues 'XX',150-153,'X',155,'X',157-159,486-492,'X',494-498,'X',500,'X',628-641,'X',643 #label VAN

REFERENCE
#authors Maury, C.P.J.
#journal J. Clin. Invest. (1991) 87:1195-1199
#title Gelsolin-related amyloidosis. Identification of the amyloid protein in Finnish hereditary amyloidosis as a fragment of variant gelsolin.
#cross-references MUID:91185597
#accession A61263
#molecule_type protein
#residues 200-213,'N',215-270 #label MAU
#experimental_source familial amyloid polyneuropathy (Finnish-type) heart and kidney amyloid fibrils
#note the substitution of Asn for 214-Asp causes disease
#authors Haltia, M.; Prelli, F.; Ghiso, J.; Kiuru, S.; Somer, H.; Palo, J.; Frangione, B.
#journal Biochem. Biophys. Res. Commun. (1990) 167:927-932
#title Amyloid protein in familial amyloidosis (Finnish type) is homologous to gelsolin, an actin-binding protein.
#cross-references MUID:90211339
#accession A34562
#molecule_type protein
#residues 200-214 #label HAL
#experimental_source diseased kidney, familial amyloidosis (Finnish type)

REFERENCE
#authors Lind, S.E.; Janmey, P.A.
#journal J. Biol. Chem. (1984) 259:13262-13266
#title Human plasma gelsolin binds to fibronectin.
#cross-references MUID:85030446
#contents annotation; fibronectin binding
#accession A34137
#authors Maury, C.P.J.; Alli, K.; Baumann, M.
#journal FEBS Lett. (1990) 260:85-87
#title Finnish hereditary amyloidosis. Amino acid sequence homology between the amyloid fibril protein and human plasma gelsoline.
#cross-references MUID:90127414

#accession A34137
#molecule_type protein
#residues 235-237;240-252;258-269 #label MA2
#experimental_source kidney, Finnish hereditary amyloidosis patient
#comment Gelsolin is a calcium-regulated, actin-modulating protein that binds to the plus (or barbed) ends of actin monomers or filaments, preventing monomer exchange (end-blocking or capping). It can promote the assembly of monomers into filaments (nucleation) as well as sever filaments already formed. A single gene encodes two forms of gelsolin; one remains associated with the cell cytoplasm, the other is secreted. In addition to its role in actin regulation, this protein is known to bind with high affinity to fibronectin.

GENETICS
#gene GDB:GSN
#cross-references GDB:120019; OMIM:105120; OMIM:137350
#map_position 9q33-9q33
#introns 48/3; 117/1
CLASSIFICATION
#superfamily gelsolin; gelsolin repeat homology
#keywords actin binding; alternative splicing; amyloid; calcium; duplication
FEATURE
1-27 #domain signal sequence #status predicted #label SIG\
28-782 #product gelsolin, plasma #status experimental #label MAT1\
51-176 #domain actin-severing #status predicted #label SEV\
53-782 #product gelsolin, cytosolic #status experimental #label MAT2\
66-397 #domain gelsolin repeat homology #label GEL1\
123-126 #region actin-actin interfilament contact\
434-782 #domain calcium-sensitive, actin binding #status predicted #label ACT\
445-764 #domain gelsolin repeat homology #label GEL2
SUMMARY
#length 782 #molecular-weight 85697 #checksum 6638

Query Match 3.3%; Score 8; DB 1; Length 782;
Best Local Similarity 100.0%; Pred. No. 3.57e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 ALLCALS 16
| | | | | | | |
Qy 17 ALLCALS 24

RESULT 5
ENTRY I48349 #type fragment
TITLE fibronectin - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 12-Jul-1996
ACCESSIONS I48349; S33445
REFERENCE I48349
#authors Polly, P.; Nicholson, R.C.
#journal Gene (1993) 137:353-354
#title Sequence of the mouse fibronectin-encoding gene promoter region.
#cross-references MUID:94131313
#accession I48349
#status preliminary; translated from GE/EMBL/DBJ
#molecule_type DNA
#residues 1-28 #label RES
#cross-references EMBL:222729; NID:g297911; PID:g297912
REFERENCE S33445
#authors Polly, P.; Nicholson, R.C.
#submission submitted to the EMBL Data Library, May 1993
#description Nucleotide sequence of the murine fibronectin gene promoter region.
#accession S33445
#status preliminary
#molecule_type DNA
#residues 1-27 #label POL
#cross-references EMBL:222729
CLASSIFICATION #superfamily fibronectin; fibronectin type I repeat homology;
```

fibronectin type II repeat homology; fibronectin type III repeat homology  
#length 28 #checksum 665

Query Match 2.9%; Score 7; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 2.50e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 GPGRLLL 12  
|||||||  
QY 35 GPGRLLL 41

RESULT 6  
ENTRY B25429 #type fragment  
TITLE T-cell receptor beta chain precursor V region (SJL73) - mouse (fragment)  
ORGANISM Mus musculus #common\_name house mouse  
DATE 02-Jun-1988 #sequence\_revision 02-Jun-1988 #text\_change 30-May-1997

ACCESSIONS B25429  
REFERENCE A94132  
#authors Behlke, M.A.; Chou, H.S.; Huppi, K.; Loh, D.Y.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:767-771  
#title Murine T-cell receptor mutants with deletions of beta-chain variable region genes.  
#cross-references MUID:86121021  
#accession B25429

##molecule\_type mRNA  
##residues 1-120 #label BEH

CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology

KEYWORDS T-cell receptor

FEATURE 11-120  
#product T-cell receptor beta chain V region SJL73  
#status predicted #label MAT

SUMMARY #length 120 #checksum 3746

Query Match 2.9%; Score 7; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 2.50e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 GPGCGPG 11  
|||||||  
QY 31 GPGCGPG 37

RESULT 7  
ENTRY A71173 #type complete  
TITLE hypothetical protein PH0581 - Pyrococcus horikoshii  
ORGANISM #formal\_name Pyrococcus horikoshii  
DATE 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Nov-1998

ACCESSIONS A71173  
REFERENCE A71000  
#authors Kavarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuwa, H.; Kikuchi, H.

#journal DNA Res. (1998) 5:53-76  
#title Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.

#cross-references MUID:98344137

#accession A71173  
#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule\_type DNA  
##residues 1-133 #label KAW

##cross-references GB:AF000002; NID:g3236129; PID:d1030613; PID:g3256987  
##experimental\_source strain OT3

fibronectin type II repeat homology; fibronectin type III repeat homology  
#length 28 #checksum 665

Query Match 2.9%; Score 7; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 2.50e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 GPGRLLL 12  
|||||||  
QY 35 GPGRLLL 41

RESULT 6  
ENTRY B25429 #type fragment  
TITLE T-cell receptor beta chain precursor V region (SJL73) - mouse (fragment)  
ORGANISM Mus musculus #common\_name house mouse  
DATE 02-Jun-1988 #sequence\_revision 02-Jun-1988 #text\_change 30-May-1997

ACCESSIONS B25429  
REFERENCE A94132  
#authors Behlke, M.A.; Chou, H.S.; Huppi, K.; Loh, D.Y.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:767-771  
#title Murine T-cell receptor mutants with deletions of beta-chain variable region genes.  
#cross-references MUID:86121021  
#accession B25429

##molecule\_type mRNA  
##residues 1-120 #label BEH

CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology

KEYWORDS T-cell receptor

FEATURE 11-120  
#product T-cell receptor beta chain V region SJL73  
#status predicted #label MAT

SUMMARY #length 120 #checksum 3746

Query Match 2.9%; Score 7; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 2.50e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 GPGCGPG 11  
|||||||  
QY 31 GPGCGPG 37

RESULT 7  
ENTRY A71173 #type complete  
TITLE hypothetical protein PH0581 - Pyrococcus horikoshii  
ORGANISM #formal\_name Pyrococcus horikoshii  
DATE 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Nov-1998

ACCESSIONS A71173  
REFERENCE A71000  
#authors Kavarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuwa, H.; Kikuchi, H.

#journal DNA Res. (1998) 5:53-76  
#title Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.

#cross-references MUID:98344137

#accession A71173  
#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule\_type DNA  
##residues 1-133 #label KAW

##cross-references GB:AF000002; NID:g3236129; PID:d1030613; PID:g3256987  
##experimental\_source strain OT3

##note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS PH0581  
#superfamily hypothetical protein MJ1081  
CLASSIFICATION #length 133 #molecular-weight 15160 #checksum 9540  
SUMMARY

Query Match 2.9%; Score 7; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 2.50e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 103 GRLLGT 109  
|||||||  
QY 37 GRLLGT 43

RESULT 8  
ENTRY UTCAB #type complete  
TITLE gonadotropin beta chain precursor - common carp  
ORGANISM #formal\_name Cyprinus carpio #common\_name common carp  
DATE 31-Aug-1979 #sequence\_revision 16-Feb-1996 #text\_change 05-Sep-1997

ACCESSIONS S29677; JT0462; S29678; A01504  
REFERENCE S29677  
#authors Chang, Y.S.; Huang, F.L.; Lo, T.B.  
#journal submitted to the EMBL Data Library, May 1991  
#title S29677  
#accession  
#status preliminary  
##molecule\_type DNA  
##residues 1-144 #label CH1  
##cross-references EMBL:X59888; NID:g62619; PID:g62620

REFERENCE JK0024  
#authors Chang, Y.S.; Huang, C.J.; Huang, F.L.; Lo, T.B.  
#journal Int. J. Pept. Protein Res. (1988) 32:556-564  
#title Primary structures of carp gonadotropin subunits deduced from cDNA nucleotide sequences.  
#cross-references MUID:89233593  
#accession JT0462

##molecule\_type mRNA  
##residues 1-144 #label CH2

REFERENCE S29677  
#authors Chang, Y.S.; Huang, F.L.; Lo, T.B.  
#journal submitted to the EMBL Data Library, May 1991  
#title S29678  
#accession  
#status preliminary  
##molecule\_type DNA  
##residues 1-144 #label CHA  
##cross-references EMBL:X59889; NID:g62621; PID:g62622

REFERENCE A90673  
#authors Jolles, J.; Burzawa-Gerard, E.; Fontaine, Y.A.; Jolles, P.  
#journal Biochimie (1977) 59:893-898  
#title The evolution of gonadotropins: some molecular data concerning a non-mammalian pituitary gonadotropin, the hormone from a teleost fish (Cyprinus carpio L.).  
#cross-references MUID:78124308  
#accession A01504

##molecule\_type protein  
##residues 28-36,'X',38-53;141-142 #label JOL

GENETICS 6/2; 65/3  
#intron superfamily pituitary glycoprotein hormone beta chain  
CLASSIFICATION #superfamily pituitary glycoprotein; pituitary  
KEYWORDS  
FEATURE 1-27  
1-27 #domain signal sequence #status predicted #label SIG  
28-117 #product gonadotropin beta chain #status predicted #label MAT

33-58,47-81,50-112,  
62-134,96-124,  
114-117  
37  
#disulfide\_bonds #status predicted  
#binding\_site carbohydrate (Asn) (covalent) #status predicted  
SUMMARY #length 144 #molecular-weight 16039 #checksum 1456

```

Query Match      2.9%; Score 7; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.50e+00;
Matches          7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 VVLLAVA 25
Qy 168 VVLLAVA 174
|||||

RESULT 9
ENTRY #type complete
TITLE N-acetyl-muramoyl-L-alanine amidase homolog yqii - Bacillus
ORGANISM subtilis
DATE #formal name Bacillus subtilis
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
E69961
REFERENCE #accession
A69580 #authors
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Prescecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takenaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references E69961
#accession E69961
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-206 #label RUN
#cross-references GB:299116; GB:AL009126; NID:g2634723; PID:e1185687;
PID:g2634853
#experimental_source strain 168

GENETICS
#gene yqii
SUMMARY #length 206 #molecular-weight 22235 #checksum 3430

Query Match      2.9%; Score 7; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 2.50e+00;
Matches          7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 ALCGLAL 15
Qy 12 ALCGLAL 18
|||||

RESULT 10
ENTRY #type fragment
TITLE CD1b protein - sheep (fragment)
ORGANISM #formal name Ovis orientalis aries, Ovis ammon aries
DATE #common_name domestic sheep
08-Sep-1997 #sequence_revision 01-Mar-1996 #text_change
S58353
REFERENCE #accession
S58353 #authors
Ferguson, E.D.; Dutia, B.M.; Hein, W.; Hopkins, J.
#submission submitted to the EMBL Data Library, July 1995
#description The ovine CD1 gene family contains at least four CD1B
homologues.
S58353
#accession S58353
#status preliminary
#molecule_type mRNA
#residues 1-232 #label FER
#cross-references EMBL:X90567; NID:g945010; PID:g945011
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 232 #checksum 6685

Query Match      2.9%; Score 7; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 2.50e+00;
Matches          7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 115 GPGRLLL 121
Qy 35 GPGRLLL 41
|||||

RESULT 11
ENTRY #type complete
TITLE thiamin biosynthesis homolog yjbt - Bacillus subtilis
ORGANISM #formal name Bacillus subtilis
DATE #cross-references E69961
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
D69845
REFERENCE #accession
A69580 #authors
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Prescecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takenaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references E69961
#accession E69961
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-206 #label RUN
#cross-references GB:299116; GB:AL009126; NID:g2634723; PID:e1185687;
PID:g2634853
#experimental_source strain 168

GENETICS
#gene yqii
SUMMARY #length 206 #molecular-weight 22235 #checksum 3430

```

Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;  
 Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;  
 Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;  
 Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,  
 K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;  
 Yoshikawa, H.; Danchin, A.  
 Nature (1997) 390:249-256  
 The complete genome sequence of the Gram-positive bacterium  
 Bacillus subtilis.  
 #journal Nature (1997) 390:249-256  
 #title The complete genome sequence of the Gram-positive bacterium  
 #cross-references MUID:98044033  
 #accession D69845  
 ##status preliminary; nucleic acid sequence not shown;  
 translation not shown  
 ##molecule\_type DNA  
 ##residues 1-256 #label KUN  
 ##cross-references GB:499110; GB:AL009126; NID:92633472; PID:ell183189;  
 PID:g2633523  
 ##experimental\_source strain 168  
 GENETICS  
 #gene yjbt  
 #classification #superfamily thiamine biosynthesis protein thiG  
 #summary #length 256 #molecular-weight 27022 #checksum 9980  
 Query Match 2.9%; Score 7; DB 2; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 2.50e+00;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 14 RLLGTG 20  
 |||||  
 QY 38 RLLGTG 44  
 RESULT 12  
 ENTRY B42424 #type fragment  
 TITLE chitinase (EC 3.2.1.14) precursor - maize (fragment)  
 ORGANISM #formal\_name Zea mays #common\_name maize  
 DATE 12-Mar-1993 #sequence\_revision 03-Feb-1994 #text\_change  
 10-Jul-1998  
 ACCESSIONS B42424  
 REFERENCE A42424  
 #authors Huynh, O.K.; Hironaka, C.M.; Levine, E.B.; Smith, C.E.;  
 Borgmeyer, J.R.; Shah, D.M.  
 #journal J. Biol. Chem. (1992) 267:6635-6640  
 #title Antifungal proteins from plants. Purification, molecular  
 cloning, and antifungal properties of chitinases from maize  
 seed.  
 #cross-references MUID:92202208  
 #accession B42424  
 ##status preliminary  
 ##molecule\_type mRNA  
 ##residues 1-268 #label HUY  
 ##note the authors translated the codon AAG for residue 104 as  
 Asn, GGC for residue 105 as Lys, TAC for residue 106  
 as Gly, GGC for residue 191 as Phe, GGC for residue  
 194 as Phe, ATG for residue 230 as Ile, CAA for  
 residue 231 as Asn and CGC for residue 232 as Gly  
 #classification #superfamily lectin-related plant chitinase; hevein  
 chitin-binding domain homology; plant chitinase homology  
 glycosidase; hydrolase; polysaccharide degradation  
 #domain hevein chitin-binding domain homology #label  
 HCB\N  
 #domain plant chitinase homology #label PCH  
 #length 268 #checksum 3749  
 SUMMARY  
 Query Match 2.9%; Score 7; DB 2; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 2.50e+00;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 6 GLALLCA 12  
 |||||  
 QY 15 GLALLCA 21

RESULT 13  
 ENTRY B45887 #type fragment  
 TITLE leukocyte differentiation antigen CD1 homolog precursor -  
 cottontail rabbit (fragment)  
 ORGANISM #formal\_name Sylvilagus sp. #common\_name cottontail rabbit  
 DATE 19-Mar-1997 #sequence\_revision 19-Dec-1997 #text\_change  
 26-Feb-1998  
 ACCESSIONS B45887  
 REFERENCE A45887  
 #authors Calabi, F.; Belt, K.T.; Yu, C.Y.; Bradbury, A.; Mandy, W.J.;  
 Milstein, C.  
 #journal Immunogenetics (1989) 30:370-377  
 #title The rabbit CD1 and the evolutionary conservation of the CD1  
 gene family.  
 #cross-references MUID:90035137  
 #accession B45887  
 ##status preliminary  
 ##molecule\_type DNA  
 ##residues 1-274 #label CAL  
 ##cross-references GB:M26249  
 #classification #superfamily class I histocompatibility antigen;  
 immunoglobulin homology  
 #summary #length 274 #checksum 1752  
 Query Match 2.9%; Score 7; DB 2; Length 274;  
 Best Local Similarity 100.0%; Pred. No. 2.50e+00;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 195 GPGRLLL 201  
 |||||  
 QY 35 GPGRLLL 41  
 RESULT 14  
 ENTRY S07532 #type complete  
 TITLE Puff II/9-1 protein precursor - fungus gnat (Sciara  
 coprophila)  
 ORGANISM #formal\_name Sciara coprophila  
 DATE 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change  
 29-Jan-1999  
 ACCESSIONS S07532  
 REFERENCE S07532  
 #authors DiBartolomeis, S.M.; Gerbi, S.A.  
 #journal J. Mol. Biol. (1989) 210:531-540  
 #title Molecular characterization of DNA puff II/9A genes in Sciara  
 coprophila.  
 #cross-references MUID:90133907  
 #accession S07532  
 ##status not compared with conceptual translation  
 ##molecule\_type DNA  
 ##residues 1-286 #label DIB  
 ##cross-references GB:X51680; NID:g10112; PID:e249512; PID:g1405811  
 GENETICS  
 #map\_position II/9A  
 #keywords coiled coil; glycoprotein  
 #feature 1-20  
 21-286  
 #domain signal sequence #status predicted #label STC\N  
 #product puff II/9-1 protein #status predicted #label  
 MAP\N  
 #binding\_site carbohydrate (Asn) (covalent) #status  
 predicted  
 SUMMARY #length 286 #molecular-weight 32034 #checksum 2862  
 Query Match 2.9%; Score 7; DB 2; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 2.50e+00;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 7 LTVVLLA 13  
 |||||  
 QY 166 LTVVLLA 172

```

RESULT 15
ENTRY
TITLE
ORGANISM S07533 #type complete
DATE puff II/9A-2 protein precursor - fungus gnat (Sciara
coprophila)
#formal_name Sciara coprophila
30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
29-Jan-1999
ACCESSIONS S07533
REFERENCE S07532
#authors DiBartolomeis, S.M.; Gerbi, S.A.
#journal J. Mol. Biol. (1989) 210:531-540
#title Molecular characterization of DNA puff II/9A genes in Sciara
coprophila.
#cross-references MUID:90133907
#accession S07533
#status not compared with conceptual translation
#molecule_type DNA
#residues 1-286 #label DIB
#cross-references GB:X51679; NID:g10113; PID:e249526; PID:g1405812
GENETICS
#map_position II/9A
#map_position colled coil; glycoprotein
KEYWORDS
FEATURE
1-20 #domain signal sequence #status predicted #label SIG\
21-286 #product puff II/9A protein #status predicted #label
MAT\
156 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 286 #molecular-weight 32621 #checksum 7272
Query Match 2.9%; Score 7; DB 2: Length 286;
Best Local Similarity 100.0%; Pred. No. 2.50e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 7 LTVVLLA 13
| | | | |
QY 166 LTVVLLA 172
Search completed: Mon Jul 12 12:13:23 1999
Job time : 22 secs.

```

\*\*\*\*\*  
WATERMAN  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
On: Mon Jul 12 12:13:44 1999; MasPar time 8.36 Seconds  
814,924 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-911-423-4  
Description: (1-241) from US08911423.pep  
Perfect Score: 241  
Sequence: 1 MAQHGAMGAFRALGGLALIC.....EEGERSAEKRGGLDW 241

Scoring table:  
TABLE unitprotatable  
Gap 60

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 3.545; Variance 0.437; scale 8.117

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match						
1	8	3.3	411	1	VGLM_HSVBC	GLYCOPROTEIN M.	1.28e-02	
2	8	3.3	782	1	GELS_HUMAN	GLYCOSOLIN PRECURSOR,	PL	1.28e-02
3	7	2.9	62	1	Y011_BACSU	HYPOTHETICAL 6.7 KD PR		1.17e+00
4	7	2.9	144	1	GTHB_CYPCA	GONADOTROPIN BETA CHAI		1.17e+00
5	7	2.9	269	1	CHTB_WAIZE	ENDOCHITINASE B PRECUR		1.17e+00
6	7	2.9	274	1	CD1_SYLFL	T-CELL SURFACE GLYCOPR		1.17e+00
7	7	2.9	286	1	P092_SCICO	PUFF II/9-2 PROTEIN PR		1.17e+00
8	7	2.9	286	1	P091_SCICO	PUFF II/9-1 PROTEIN PR		1.17e+00
9	7	2.9	297	1	SYN4_HUMAN	SYNTAXIN 4.		1.17e+00
10	7	2.9	324	1	MAS_MOUSE	MAS PROTO-ONCOGENE.		1.17e+00
11	7	2.9	335	1	CD1D_HUMAN	T-CELL SURFACE GLYCOPR		1.17e+00
12	7	2.9	359	1	WNSA_AMBME	WNT-5A PROTEIN PRECURS		1.17e+00
13	7	2.9	401	1	PILC_PSEPU	FIBRILLAR ASSEMBLY PROT		1.17e+00
14	7	2.9	508	1	VGLG_THNV	SPIKE GLYCOPROTEIN PRE		1.17e+00
15	7	2.9	524	1	VGLG_CHAV	SPIKE GLYCOPROTEIN PRE		1.17e+00
16	7	2.9	555	1	NUSM_CANPA	NADH-UBIQUINONE OXIDOR		1.17e+00
17	7	2.9	598	1	NUSM_BRALA	NADH-UBIQUINONE OXIDOR		1.17e+00
18	7	2.9	600	1	NUSM_METSE	NADH-UBIQUINONE OXIDOR		1.17e+00
19	7	2.9	603	1	NUSM_GORGO	NADH-UBIQUINONE OXIDOR		1.17e+00
20	7	2.9	603	1	NUSM_PONPA	NADH-UBIQUINONE OXIDOR		1.17e+00
21	7	2.9	603	1	NUSM_PANPA	NADH-UBIQUINONE OXIDOR		1.17e+00
22	7	2.9	603	1	NUSM_HUMAN	NADH-UBIQUINONE OXIDOR		1.17e+00
23	7	2.9	603	1	NUSM_PONPY	NADH-UBIQUINONE OXIDOR		1.17e+00

24	7	2.9	603	1	NUSM_HYLLA	NADH-UBIQUINONE OXIDOR	1.17e+00
25	7	2.9	603	1	NUSM_PANTR	NADH-UBIQUINONE OXIDOR	1.17e+00
26	7	2.9	641	1	NUSM_ALMA	CYTCHROME C-TYPE BIOS	1.17e+00
27	7	2.9	653	1	CCMF_RHOCA	NADH-UBIQUINONE OXIDOR	1.17e+00
28	7	2.9	666	1	NUSM_CHOCR	T-CADHERIN PRECURSOR	1.17e+00
29	7	2.9	713	1	CADD_HUMAN	HYPOTHETICAL 85.3 KD P	1.17e+00
30	7	2.9	750	1	YF03_MYCPN	CARTILAGE OLIGOMERIC M	1.17e+00
31	7	2.9	757	1	COMP_HUMAN	HYPOTHETICAL 94.9 KD P	1.17e+00
32	7	2.9	830	1	YJG2_YEAST	CHROMOSOME CONDENSATIO	1.17e+00
33	7	2.9	1469	1	DP27_CAEEL	FIBRONECTIN PRECURSOR	1.17e+00
34	7	2.9	2477	1	FINC_MOUSE	FIBRONECTIN PRECURSOR	1.17e+00
35	7	2.9	2477	1	FINC_RAT	PROTEIN F9 HOMOLOG (FR	6.29e+01
36	6	2.5	73	1	VF09_ORFN2	UV-INDUCED PROTEIN UVI	6.29e+01
37	6	2.5	105	1	UV15_SCHPO	NONSTRUCTURAL PROTEIN	6.29e+01
38	6	2.5	118	1	VNS2_IAMYN	NONSTRUCTURAL PROTEIN	6.29e+01
39	6	2.5	121	1	VNS2_IAP10	NONSTRUCTURAL PROTEIN	6.29e+01
40	6	2.5	121	1	VNS2_IAP11	NONSTRUCTURAL PROTEIN	6.29e+01
41	6	2.5	121	1	VNS2_IAP12	NONSTRUCTURAL PROTEIN	6.29e+01
42	6	2.5	121	1	VNS2_IAP18	NONSTRUCTURAL PROTEIN	6.29e+01
43	6	2.5	121	1	VNS2_IAM6	NONSTRUCTURAL PROTEIN	6.29e+01
44	6	2.5	121	1	VNS2_IAMAN	NONSTRUCTURAL PROTEIN	6.29e+01
45	6	2.5	121	1	VNS2_IATKB	NONSTRUCTURAL PROTEIN	6.29e+01
46	6	2.5	121	1	VNS2_IATKR	NONSTRUCTURAL PROTEIN	6.29e+01
47	6	2.5	121	1	VNS2_IANA	NONSTRUCTURAL PROTEIN	6.29e+01
48	6	2.5	121	1	VNS2_IAP12	NONSTRUCTURAL PROTEIN	6.29e+01
49	6	2.5	135	1	YOR6_ADEG1	HYPOTHETICAL 15.1 KD P	6.29e+01
50	6	2.5	145	1	URL_CYPCA	UROTENSIN I PRECURSOR	6.29e+01
51	6	2.5	167	1	YMW1_YEAST	HYPOTHETICAL 18.7 KD P	6.29e+01
52	6	2.5	173	1	YPII_LACDL	HYPOTHETICAL 19.8 KD P	6.29e+01
53	6	2.5	178	1	LACB_PIG	BETA-LACTOGLOBULIN IA	6.29e+01
54	6	2.5	216	1	YXBE_BACSU	HYPOTHETICAL 23.1 KD P	6.29e+01
55	6	2.5	220	1	TPIS_CHLAU	TRIOSEPHOSPHATE ISOMER	6.29e+01
56	6	2.5	221	1	YSLC_YERPS	YOP PROTEINS TRANSLOCA	6.29e+01
57	6	2.5	223	1	YSLC_YEREN	YOP PROTEINS TRANSLOCA	6.29e+01
58	6	2.5	236	1	VRP4_SALTY	VIRULENCE PROTEIN MKFA	6.29e+01
59	6	2.5	240	1	VRP3_SALTY	27.5 KD VIRULENCE PROT	6.29e+01
60	6	2.5	247	1	YV35_MYCLE	HYPOTHETICAL 27.0 KD P	6.29e+01
61	6	2.5	269	1	YIB1_YEAST	HYPOTHETICAL 26.3 KD P	6.29e+01
62	6	2.5	283	1	VAM3_YEAST	SYNTAXIN-RELATED PROTE	6.29e+01
63	6	2.5	310	1	BCHG_CHLAU	BACTERIOCHLOROPHYLL SY	6.29e+01
64	6	2.5	331	1	TGL_STRSS	PROTEIN-GLUTAMINE GAMM	6.29e+01
65	6	2.5	348	1	UN97_CAEEL	HOMEOBOX PROTEIN UNC-9	6.29e+01
66	6	2.5	353	1	GBQ_CANFA	GUANINE NUCLEOTIDE-BIN	6.29e+01
67	6	2.5	353	1	GBQ_MOUSE	GUANINE NUCLEOTIDE-BIN	6.29e+01
68	6	2.5	353	1	GBQ_PAITY	GUANINE NUCLEOTIDE-BIN	6.29e+01
69	6	2.5	355	1	GB14_MOUSE	GUANINE NUCLEOTIDE-BIN	6.29e+01
70	6	2.5	359	1	GB11_MOUSE	GUANINE NUCLEOTIDE-BIN	6.29e+01
71	6	2.5	359	1	GB11_XENLA	GUANINE NUCLEOTIDE-BIN	6.29e+01
72	6	2.5	359	1	GB11_HUMAN	GUANINE NUCLEOTIDE-BIN	6.29e+01
73	6	2.5	368	1	RIBD_SYNY3	RIBOFLAVIN-SPECIFIC DE	6.29e+01
74	6	2.5	370	1	YI81_ECOLI	INSERTION ELEMENT IS18	6.29e+01
75	6	2.5	371	1	YI44_ECOLI	INSERTION ELEMENT IS42	6.29e+01
76	6	2.5	375	1	YI31_ONCMY	HYPOTHETICAL PROTEIN T	6.29e+01
77	6	2.5	424	1	ZP3A_CALSQ	ZONA PELLUCIDA SPERM-B	6.29e+01
78	6	2.5	425	1	YQCE_ECOLI	HYPOTHETICAL 46.8 KD P	6.29e+01
79	6	2.5	454	1	YUAF_CAEEL	PUTATIVE SERINE CARBOX	6.29e+01
80	6	2.5	490	1	YHAF_ECOLI	HYPOTHETICAL 54.2 KD P	6.29e+01
81	6	2.5	490	1	YHR3_YEAST	HYPOTHETICAL 54.2 KD P	6.29e+01
82	6	2.5	507	1	YV45_CAEEL	HYPOTHETICAL 55.1 KD P	6.29e+01
83	6	2.5	547	1	ICA3_HUMAN	INTERCELLULAR ADHESION	6.29e+01
84	6	2.5	590	1	VG28_HSV11	HYPOTHETICAL GENE 28 P	6.29e+01
85	6	2.5	658	1	CTR2_HUMAN	LOW-AFFINITY CATIONIC	6.29e+01
86	6	2.5	685	1	YVCA_BACSU	HYPOTHETICAL 73.6 KD P	6.29e+01
87	6	2.5	703	1	Y411_RHISN	HYPOTHETICAL 76.2 KD P	6.29e+01
88	6	2.5	707	1	SR72_SCHMA	SIGNAL RECOGNITION PAR	6.29e+01
89	6	2.5	724	1	RRPO_THVD	RNA-DIRECTED RNA POLYM	6.29e+01
90	6	2.5	754	1	BGLB_CLOTM	THERMOSTABLE BETA-GLUC	6.29e+01
91	6	2.5	774	1	RRPO_TCV	PROBABLY RNA-DIRECTED	6.29e+01
92	6	2.5	808	1	SYFB_SYNP7	PHENYLALANYL-TRNA SYNT	6.29e+01
93	6	2.5	870	1	YCSQ_YEAST	HYPOTHETICAL 96.1 KD P	6.29e+01
94	6	2.5	1065	1	VINC_HUMAN	VINCULIN	6.29e+01
95	6	2.5	1065	1	VINC_MOUSE	VINCULIN	6.29e+01
96	6	2.5	1354	1	VGR2_HUMAN	VASCULAR ENDOTHELIAL G	6.29e+01

97 6 2.5 1938 1 MYSD\_CABEL MYOSIN HEAVY CHAIN D ( 6.29e+01  
98 6 2.5 2090 1 HFC1\_MESAU HOST CELL FACTOR C1 (H 6.29e+01  
99 6 2.5 3988 1 POLG\_BVDV GENOME POLYPROTEIN. 6.29e+01  
100 6 2.5 3988 1 POLG\_BVDV GENOME POLYPROTEIN. 6.29e+01

## ALIGNMENTS

RESULT 1  
ID VCLM\_HSVBC STANDARD; PRT; 411 AA.  
AC P52370;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE GLYCOPROTEIN M.  
GN CM OR UL10.  
OS BOVINE HERPESVIRUS TYPE 1 (STRAIN COOPER).  
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;  
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.  
[1]  
SEQUENCE FROM N.A.  
MEDLINE; 95313343.  
RA VICEK C., BENES V., LU Z., KUTISH G.F., PACES V., ROCK D.,  
RA LETCHWORTH G.J., SCHWYZER M.;  
RT "Nucleotide sequence analysis of a 30-kb region of the bovine  
RT herpesvirus 1 genome which exhibits a colinear gene arrangement with  
RT the UL21 to UL4 genes of herpes simplex virus.";  
RL VIROLOGY 210:100-108(1995).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: TO OTHER HERPESVIRUS GLYCOPROTEIN M.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
-----  
DR EMBL; Z48053; G971324; -  
KW TRANSMEMBRANE; GLYCOPROTEIN.  
FT TRANSWMEM 14 34 POTENTIAL.  
FT TRANSWMEM 89 109 POTENTIAL.  
FT TRANSWMEM 133 153 POTENTIAL.  
FT TRANSWMEM 159 179 POTENTIAL.  
FT TRANSWMEM 213 233 POTENTIAL.  
FT TRANSWMEM 249 269 POTENTIAL.  
FT TRANSWMEM 277 297 POTENTIAL.  
FT TRANSWMEM 318 338 POTENTIAL.  
FT CARBOHYD 57 57 POTENTIAL.  
SQ SEQUENCE 411 AA; 43029 MW; E662D006 CRC32;

Query Match 3.3%; Score 8; DB 1; Length 411;  
Best Local Similarity 100.0%; Pred. No. 1.28e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 93 VVLA VAA 100  
|||||||  
QY 168 VVLA VAA 175

RESULT 2  
ID GELS\_HUMAN STANDARD; PRT; 782 AA.  
AC P06396;  
DT 01-JAN-1988 (REL. 06, CREATED)  
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE GELSOLIN PRECURSOR, PLASMA (ACTIN-DEPOLYMERIZING FACTOR) (ADF)  
DE (BREVIN) (AGEL).  
GN GSN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87014807.  
RA KWIAKOWSKI D.J., STOSSEL T.P., ORKIN S.H., MOLE J.E., COLTEN H.R.,  
RA YIN H.L.;  
RT "Plasma and cytoplasmic gelsolins are encoded by a single gene and  
RT contain a duplicated actin-binding domain.";  
RL NATURE 323:455-458(1986).  
RN  
RP FIBRONECTIN BINDING.  
RX MEDLINE; 85030446.  
RA LIND S.E., JANNEY P.A.;  
RT "Human plasma gelsolin binds to fibronectin.";  
RL J. BIOL. CHEM. 259:13262-13266(1984).  
RN  
RP  
RP IDENTITY OF FAF AMYLOID PROTEIN WITH GELSOLIN.  
RX MEDLINE; 90211339.  
RA HALTIA M., PRELLI F., GHISO J., KIURU S., SOMMER H., PALO J.,  
RA FRANGIONE B.;  
RT "Amyloid protein in familial amyloidosis (Finnish type) is homologous  
RT to gelsolin, an actin-binding protein.";  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 167:927-932(1990).  
RN  
RP IDENTITY OF FAF AMYLOID PROTEIN WITH GELSOLIN.  
RX MEDLINE; 90127414.  
RA MAURY C.P.J., ALLI K., BAUMANN M.;  
RT "Finnish hereditary amyloidosis. Amino acid sequence homology between  
RT the amyloid fibril protein and human plasma gelsoline.";  
RL FEBS LETT. 260:85-87(1990).  
RN  
RP DISULFIDE BOND.  
RX MEDLINE; 96312883.  
RA WEN D., CORINA K., CHOW E.P., MILLER S., JANNEY P.A., PEPINSKY R.B.;  
RT "The plasma and cytoplasmic forms of human gelsolin differ in  
RT disulfide structure.";  
RL BIOCHEMISTRY 35:9700-9709(1996).  
RN  
RP DISULFIDE BOND.  
RX MEDLINE; 97157527.  
RA ALLEN P.G.;  
RT "Functional consequences of disulfide bond formation in gelsolin.";  
RL FEBS LETT. 401:89-94(1997).  
RN  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 28-503.  
RX MEDLINE; 93361115.  
RA MCLAUGHLIN P.J., GOOCH J.T., MANNHERZ H.-G., WEEDS A.G.;  
RT "Structure of gelsolin segment 1-actin complex and the mechanism of  
RT filament severing.";  
RL NATURE 364:685-692(1993).  
RN  
RP STRUCTURE BY NMR OF 177-196.  
RX MEDLINE; 96167246.  
RA XIAN W., VEGNERS R., JANNEY P.A., BRAUNLIN W.H.;  
RT "Spectroscopic studies of a phosphoinositide-binding peptide from  
RT gelsolin: behavior in solutions of mixed solvent and anionic  
RT micelles.";  
RL BIOPHYS. J. 69:2695-2702(1995).  
RN  
RP VARIANT FAF ASN-214.  
RX MEDLINE; 91097517.  
RA GHISO J., HALTIA M., PRELLI F., NOVELLO J., FRANGIONE B.;  
RT "Gelsolin variant (Asn-187) in familial amyloidosis, Finnish type.";  
RL BIOCHEM. J. 272:827-830(1990).  
RN  
RP VARIANTS FAF ASN-214 AND TYR-214.  
RX MEDLINE; 93265143.  
RA DE LA CHAPELLE A., TOLVANEN R., BOYSEN G., SANTAVY J.,  
RA BLEEKER-WAGEMAKERS L., MAURY C.P.J., KERE J.;  
RT "Gelsolin-derived familial amyloidosis caused by asparagine or  
RT tyrosine substitution for aspartic acid at residue 187.";  
RL NAT. GENET. 2:157-160(1992).  
CC  
CC -1- FUNCTION: GELSOLIN IS A CALCIUM-REGULATED, ACTIN-MODULATING  
CC PROTEIN THAT BINDS TO THE PLUS (OR BARBED) ENDS OF ACTIN MONOMERS

OR FILAMENTS, PREVENTING MONOMER EXCHANGE (END-BLOCKING OR CAPPING). IT CAN PROMOTE THE ASSEMBLY OF MONOMERS INTO FILAMENTS (NUCLEATION) AS WELL AS SEVER FILAMENTS ALREADY FORMED.

-!- SUBCELLULAR LOCATION: SECRETED (PLASMA FORM) AND CYTOPLASMIC.

-!- TISSUE SPECIFICITY: PHAGOCYtic CELLS, PLATELETS, FIBROBLASTS, NONMUSCLE CELLS, SMOOTH AND SKELETAL MUSCLE CELLS.

-!- DISEASE: DEFECTS IN GSN ARE THE CAUSE OF FAMILIAL AMYLOIDOSIS, FINNISH TYPE (FAF), AN INHERITED FORM OF SYSTEMIC AMYLOIDOSIS CLINICALLY CHARACTERIZED BY CRANIAL NEUROPATHY AND LATTICE CORNEAL DYSTROPHY.

-!- IN ADDITION TO ITS ROLE IN ACTIN REGULATION, THIS PROTEIN IS KNOWN TO BIND WITH HIGH AFFINITY TO FIBRONECTIN.

-!- ALTERNATIVE PRODUCTS: PLASMA GELSOLIN AND CYTOPLASMIC GELSOLIN ARE DERIVED FROM A SINGLE GENE BY ALTERNATE INITIATION SITES AND DIFFERENTIAL SPLICING.

-!- SIMILARITY: BELONGS TO THE VILIN/GELSOLIN FAMILY.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; X04412; G736249; -

PIR; A03011; FAHUP

PIR; A34137; A34137.

PIR; A34562; A34562.

PDB; 1SOL; 08-MAR-96.

MIM; 137350; -

MIM; 105120; -

PFAM; PF00626; Gelsolin; 6.

ACTIN-BINDING; REPEAT; CALCIUM; ALTERNATIVE INITIATION; SIGNAL; CAPPING PROTEIN; AMYLOID; DISEASE MUTATION; 3D-STRUCTURE.

SIGNAL 1 27

CHAIN 28 782

CHAIN 53 782

DOMAIN 53 782

DOMAIN 434 782

SITE 123 126

REPEAT 57 416

DOMAIN 417 433

REPEAT 434 782

DOMAIN 57 176

REPEAT 89 97

REPEAT 112 128

REPEAT 150 161

DOMAIN 177 293

DOMAIN 210 218

REPEAT 224 240

REPEAT 263 274

DOMAIN 294 416

REPEAT 326 334

REPEAT 342 358

REPEAT 382 393

DOMAIN 434 555

REPEAT 468 476

REPEAT 490 506

REPEAT 528 539

DOMAIN 556 661

REPEAT 588 596

REPEAT 602 618

REPEAT 634 645

DOMAIN 662 782

REPEAT 693 701

REPEAT 707 723

REPEAT 749 761

DOMAIN 162 169

DOMAIN 188 196

DISULFID 215 228

-----

GELSOLIN, PLASMA.

GELSOLIN, CYTOPLASMIC.

ACTIN-SEVERING (POTENTIAL).

ACTIN-BINDING, CA-SENSITIVE (POTENTIAL).

ACTIN-ACTIN INTERFILAMENT CONTACT POINT.

1ST HALF.

HINGE REGION.

2ND HALF.

1A.

MOTIF B.

MOTIF A.

MOTIF C.

1B.

MOTIF B.

MOTIF A.

MOTIF C.

1C.

MOTIF B.

MOTIF A.

MOTIF C.

2A.

MOTIF B.

MOTIF A.

MOTIF C.

2B.

MOTIF B.

MOTIF A.

MOTIF C.

2C.

MOTIF B.

MOTIF A.

MOTIF C.

POLYPHOSPHOSINOSITIDE BINDING (BY SIMILARITY).

POLYPHOSPHOSINOSITIDE BINDING (BY SIMILARITY).

IN PLASMA FORM ONLY.

FT VARIANT 214 214 D -> N (IN FAF).

FT VARIANT 214 214 D -> Y (IN FAF).

FT VARSPLIC 1 51 MISSING (IN CYTOPLASMIC FORM).

FT INIT MET 52 52 FOR CYTOPLASMIC FORM.

SQ SEQUENCE 782 AA; 85697 MW; E03DDEBA CRC32;

Query Match 3.3%; Score 8; DB 1; Length 782;

Best Local Similarity 100.0%; Pred. No. 1.28e-02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 ALLCALSL 16

QY 17 ALLCALSL 24

-----

RESULT 3

ID YQII\_BACSU STANDARD; PRT; 62 AA.

AC P54525;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 6.7 KD PROTEIN IN SPOOA-MMGA INTERGENIC REGION.

GN YQII.

OS BACILLUS SUBTILIS.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

OC BACILLUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-168 / JH642;

RA KOBAYASHI Y., MIZUNO M., MASUDA S., TAKEMARU K., HOSONO S.,

RA SATO T., TAKEUCHI M.

RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

CC

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; D84432; D1013263; -

DR SUBTILIST; BG11717; YQII.

KW HYPOTHETICAL PROTEIN; ATP-BINDING.

FT NP\_BIND 24 31 ATP (POTENTIAL).

SQ SEQUENCE 62 AA; 6704 MW; 0003AAFO CRC32;

Query Match 2.9%; Score 7; DB 1; Length 62;

Best Local Similarity 100.0%; Pred. No. 1.17e-00;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 ALCGLAL 15

QY 12 ALCGLAL 18

-----

RESULT 4

ID GTHB\_CYPCA STANDARD; PRT; 144 AA.

AC P01235;

DT 21-JUL-1986 (REL. 01, CREATED)

DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE GONADOTROPIN BETA CHAIN PRECURSOR.

OS CYPRINUS CARPIO (COMMON CARP).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;

OC TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;

OC CYPRINIDAE; CYPRININAE; CYPRINUS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 89233593.

RA CHANG Y.S., HUANG C.-J., HUANG F.-L., LO T.-B.;

RT "Primary structures of carp gonadotropin subunits deduced from cDNA nucleotide sequences.";



OC LAGOMORPHA; LEPORIDAE; SYLVILAGUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90035137.  
RA CALABI F., BELT K.T., YU C.Y., BRADBURY A., MANDY W.J., MILSTEIN C.;  
RT "The rabbit CDI and the evolutionary conservation of the CDI gene  
family.";  
RL IMMUNOGENETICS 30:370-377(1989).  
CC  
CC -!- FUNCTION: NOT KNOWN.  
CC -!- SUBUNIT: ASSOCIATES NON-COVALENTLY WITH BETA-2-MICROGLOBULIN.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: EXPRESSED ON CORTICAL THYMOCYTES, ON CERTAIN  
CC T-CELL LEUKEMIAS, AND IN VARIOUS OTHER TISSUES.  
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M26249; E27619; -  
CC PFAM: PF00047; ig: 1.  
CC HSP: P11609; ICD1.  
CC GLYCOPROTEIN; IMMUNOGLOBULIN FOLD; T-CELL.  
FT NON\_TER 1 1  
FT DOMAIN 1 88 EXTRACELLULAR ALPHA-1.  
FT DOMAIN 89 181 EXTRACELLULAR ALPHA-2.  
FT DOMAIN 182 274 EXTRACELLULAR ALPHA-3.  
FT DISULFID 99 163 BY SIMILARITY.  
FT DISULFID 203 238 BY SIMILARITY.  
FT CARBOHYD 17 17 POTENTIAL.  
FT CARBOHYD 39 39 POTENTIAL.  
FT CARBOHYD 105 105 POTENTIAL.  
FT CARBOHYD 160 160 POTENTIAL.  
FT NON\_TER 274 274  
SQ SEQUENCE 274 AA; 30990 MW; DFE4FLED CRC32;

Query Match 2.9%; Score 7; DB 1; Length 274;  
Best Local Similarity 100.0%; Pred. No. 1.17e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 195 GPGRLLL 201  
35 GPGRLLL 41

RESULT 7  
ID PU92\_SCICO STANDARD; PRT; 286 AA.  
AC P22312;  
DT 01-AUG-1991 (REL. 19, CREATED)  
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE PU92 II/9-2 PROTEIN PRECURSOR.  
GN II/9-2.  
OS SCIARA COPROPHILA (FUNGUS GNAT).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; NEMATOCERA; SCIAROIDAE; BRADYSIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6980;  
RX MEDLINE: 90133907.  
RA DIBARTOLOMEIS S.M., GERBI S.A.;  
RT "Molecular characterization of DNA puff II/9A genes in Sciara  
coprophila.";  
RL J. MOL. BIOL. 210:531-540(1989).  
CC -!- THE DNA PUFF II/9 PROTEINS HAVE A PROPOSED INTERMOLECULAR  
CC COILED COIL STRUCTURE WITH POSSIBLY INTERMOLECULAR DISULFIDE  
CC BRIDGES FORMED BY NUMEROUS CYSTEINE RESIDUES IN POSITION D OF  
CC THE HEPTAD REPEAT.  
CC -!- SIMILARITY: 76% IDENTICAL TO THE PUFF II/9-1 PROTEIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X51680; E249512; -  
CC PIR: S07532; S07532.  
CC SIGNAL; COILED COIL; GLYCOPROTEIN.  
FT SIGNAL 1 219 OR 21 (POTENTIAL).  
FT CHAIN ?20 286 PUFF II/9-1 PROTEIN.  
FT DOMAIN 61 235 HELICAL (POTENTIAL).  
FT CARBOHYD 156 156 POTENTIAL.  
SQ SEQUENCE 286 AA; 32034 MW; 87D52C9E CRC32;

Query Match 2.9%; Score 7; DB 1; Length 286;  
Best Local Similarity 100.0%; Pred. No. 1.17e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X51679; E249526; -  
CC PIR: S07533; S07533.  
CC SIGNAL; COILED COIL; GLYCOPROTEIN.  
FT SIGNAL 1 219 OR 21 (POTENTIAL).  
FT CHAIN ?20 286 PUFF II/9-2 PROTEIN.  
FT DOMAIN 61 235 HELICAL (POTENTIAL).  
FT CARBOHYD 156 156 POTENTIAL.  
SQ SEQUENCE 286 AA; 32621 MW; F8EA569A CRC32;

Query Match 2.9%; Score 7; DB 1; Length 286;  
Best Local Similarity 100.0%; Pred. No. 1.17e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 LTVVLLA 13  
166 LTVVLLA 172

RESULT 8  
ID PU91\_SCICO STANDARD; PRT; 286 AA.  
AC P22311;  
DT 01-AUG-1991 (REL. 19, CREATED)  
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE PU91 II/9-1 PROTEIN PRECURSOR.  
GN II/9-1.  
OS SCIARA COPROPHILA (FUNGUS GNAT).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; NEMATOCERA; SCIAROIDAE; BRADYSIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6980;  
RX MEDLINE: 90133907.  
RA DIBARTOLOMEIS S.M., GERBI S.A.;  
RT "Molecular characterization of DNA puff II/9A genes in Sciara  
coprophila.";  
RL J. MOL. BIOL. 210:531-540(1989).  
CC -!- THE DNA PUFF II/9 PROTEINS HAVE A PROPOSED INTERMOLECULAR  
CC COILED COIL STRUCTURE WITH POSSIBLY INTERMOLECULAR DISULFIDE  
CC BRIDGES FORMED BY NUMEROUS CYSTEINE RESIDUES IN POSITION D OF  
CC THE HEPTAD REPEAT.  
CC -!- SIMILARITY: 76% IDENTICAL TO THE PUFF II/9-2 PROTEIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X51680; E249512; -  
CC PIR: S07532; S07532.  
CC SIGNAL; COILED COIL; GLYCOPROTEIN.  
FT SIGNAL 1 219 OR 21 (POTENTIAL).  
FT CHAIN ?20 286 PUFF II/9-1 PROTEIN.  
FT DOMAIN 61 235 HELICAL (POTENTIAL).  
FT CARBOHYD 156 156 POTENTIAL.  
SQ SEQUENCE 286 AA; 32034 MW; 87D52C9E CRC32;

Query Match 2.9%; Score 7; DB 1; Length 286;  
Best Local Similarity 100.0%; Pred. No. 1.17e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 LTVVLLA 13  
| | | | |  
Qy 166 LTVVLLA 172

Query Match 2.9%; Score 7; DB 1; Length 297;  
Best Local Similarity 100.0%; Pred. No. 1.17e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 283 TVVLLAV 289  
| | | | |  
Qy 167 TVVLLAV 173

RESULT 10  
ID MAS\_MOUSE STANDARD; PRT; 324 AA.  
AC P30554; O35944;  
DT 01-APR-1993 (REL. 25, CREATED)  
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE MAS PROTO-ONCOGENE.  
GN MAS1 OR MAS-1 OR MAS.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE=LIVER;  
RX MEDLINE; 95094925.  
RA METZGER R., BADER M., LUDWIG T., BERBERICH C., BUNNEMANN B.,  
RA GANTEN D.;  
RT "Expression of the mouse and rat mas proto-oncogene in the brain and  
peripheral tissues.";  
RL FEBS LETT. 357:27-32(1995).  
RN [2]  
RP SEQUENCE OF 1-25 FROM N.A.  
RC STRAIN-BCBA; TISSUE=TESTIS;  
RX MEDLINE; 97422605.  
RA SCHWEIFER N., VALK P.J., DELWEL R., COX R., FRANCIS F.,  
RA MEIER-EWERT S., LEHRACH H., BARLOW D.P.;  
RT "Characterization of the C3 YAC contig from proximal mouse chromosome  
17 and analysis of allelic expression of genes flanking the imprinted  
Igf2r gene.";  
RL GENOMICS 43:285-297(1997).  
CC -!- FUNCTION: NOT KNOWN.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE A RECEPTOR FOR ANGIOTENSIN  
II.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X67735; G53012; -  
EMBL; U96273; G2352058; -  
PIR; S29619; S29619.  
DR GCRDB; GCR\_0660; -  
DR GCRDB; GCR\_2448; -  
DR MGD; MGI:96918; MAS1.  
DR PROSITE; PS00237; G-PROTEIN\_RECEPTOR; 1.  
DR PFAM; PF00001; 7tm1.1.  
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;  
FT PROTO-ONCOGENE.  
FT DOMAIN 1 35 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 36 60 1 (POTENTIAL).  
FT DOMAIN 61 64 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 65 86 2 (POTENTIAL).  
FT DOMAIN 87 103 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 104 127 3 (POTENTIAL).  
FT DOMAIN 128 148 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 149 171 4 (POTENTIAL).  
FT DOMAIN 172 184 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 185 205 5 (POTENTIAL).

Db 7 LTVVLLA 13  
| | | | |  
Qy 166 LTVVLLA 172

RESULT 9  
ID SYN4\_HUMAN STANDARD; PRT; 297 AA.  
AC Q12846; Q15525;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE SYNTAXIN 4.  
GN STX4A OR STX4.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE; 94266173.  
RA LI H., HODGE D.R., PEI G.K., SETH A.;  
RT "Isolation and sequence analysis of the human syntaxin-encoding  
gene.";  
RL GENE 143:303-304(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SKETAL MUSCLE;  
RX MEDLINE; 96332494.  
RA JAGADISH M.N., FERNANDEZ C.S., HEWISH D.R., MACAULAY S.L.,  
RA GOUGH K.H., GRUSOVIN J., VERKUYEN A., COSGROVE L., ALAFACI A.,  
RA FRENKEL M.J., WARD C.W.;  
RT "Insulin-responsive tissues contain the core complex protein SNAP-25  
(synaptosomal-associated protein 25) A and B isoforms in addition to  
syntaxin 4 and synaptobrevins 1 and 2.";  
RL BIOCHEM. J. 317:945-954(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PERIPHERAL BLOOD NEUTROPHILS;  
RA NABOKINA S., LAZO P.A., MOLLINEDO F.;  
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LENS EPITHELIUM;  
RA RAE J.L., SHEPARD A.R.;  
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: POTENTIALLY INVOLVED IN DOCKING OF SYNAPTIC VESICLES  
AT PRESYNAPTIC ACTIVE ZONES.  
CC -!- SIMILARITY: BELONGS TO THE SYNTAXIN/EPIMORPHIN FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U07158; G463907; -  
EMBL; X85784; G758105; -  
DR EMBL; AJ000541; E332032; -  
DR EMBL; AF026007; G2570870; -  
DR MIM; 186591; -  
DR PROSITE; PS00914; SYNTAXIN; 1.  
DR PFAM; PF00804; Syntaxin; 1.  
DR HSSP; P32851; 1BR0.  
KW COILED COIL; TRANSMEMBRANE.  
FT DOMAIN 43 163 COILED COIL (POTENTIAL).  
FT TRANSMEM 199 222 COILED COIL (POTENTIAL).  
FT TRANSMEM 276 296 POTENTIAL.  
FT CONFLICT 174 174 E -> D (IN REF. 1).  
FT CONFLICT 269 269 A -> V (IN REF. 1).  
SQ SEQUENCE 297 AA; 34180 MW; 4A8CA175 CRC32;

FT DOMAIN 206 223 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 224 244 6 (POTENTIAL).  
FT DOMAIN 245 262 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 263 283 7 (POTENTIAL).  
FT DOMAIN 284 324 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 5 5 POTENTIAL.  
FT CARBOHYD 16 16 POTENTIAL.  
FT CARBOHYD 21 21 POTENTIAL.  
FT CONFLICT 19 19 S -> I (IN REF. 2).  
SQ SEQUENCE 324 AA; 36904 MW; 5D83B5A9 CRC32;  
  
Query Match 2.9%; Score 7; DB 1; Length 324;  
Best Local Similarity 100.0%; Pred. No. 1.17e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 153 ALLCALLS 159  
QY 17 ALLCALLS 23  
|||||||  
RESULT 11  
ID CD1D\_HUMAN STANDARD; PRT; 335 AA.  
AC F15813;  
DT 01-APR-1990 (REL. 14, CREATED)  
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE T-CELL SURFACE GLYCOPROTEIN CD1D PRECURSOR (CD1D ANTIGEN) (R3G1).  
GN CD1D.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89196496.  
RA CALABI F., JARVIS J.M., MARTIN L., MILSTEIN C.;  
RT "Two classes of CD1 genes";  
RL EUR. J. IMMUNOL. 19:285-292(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89098892.  
RA BALK S.P., BLEICHER P.A., TERHORST C.;  
RT "Isolation and characterization of a cDNA and gene coding for a  
fourth CD1 molecule";  
RL PROC. NATL. ACAD. SCI. U.S.A. 86:252-256(1989).  
CC -!- FUNCTION: NOT KNOWN.  
CC -!- SUBUNIT: ASSOCIATES NON-COVALENTLY WITH BETA-2-MICROGLOBULIN.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: EXPRESSED ON CORTICAL THYMOCYTES, ON CERTAIN  
T-CELL LEUKEMIAS, AND IN VARIOUS OTHER TISSUES.  
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD1 entry;  
WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd1.htm".  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; L38820; G619798;  
DR EMBL; L38815; G619798; JOINED.  
DR EMBL; L38817; G619798; JOINED.  
DR EMBL; L38816; G619798; JOINED.  
DR EMBL; L38818; G619798; JOINED.  
DR EMBL; L38819; G619798; JOINED.  
DR EMBL; L38819; G619798; JOINED.  
DR EMBL; X14974; G296639;  
DR EMBL; J04142; G619800;  
DR PIR; S07715; HLHUR3.  
DR PIR; A32217; A32217.  
DR MIM; 188410;  
DR PFAM; PF00047; ig; 1.

DR HSP; P11609; ICD1.  
KW GLYCOPROTEIN; SIGNAL; TRANSMEMBRANE; IMMUNOGLOBULIN FOLD; T-CELL;  
MULTIGENE FAMILY.  
FT SIGNAL 1 20  
FT CHAIN 21 335 CD1D ANTIGEN.  
FT DOMAIN 21 109 EXTRACELLULAR ALPHA-1.  
FT DOMAIN 110 202 EXTRACELLULAR ALPHA-2.  
FT DOMAIN 203 295 EXTRACELLULAR ALPHA-3.  
FT TRANSMEM 296 321 POTENTIAL.  
FT DOMAIN 322 335 CYTOPLASMIC (POTENTIAL).  
FT DISULFID 120 184 BY SIMILARITY.  
FT DISULFID 224 279 BY SIMILARITY.  
FT CARBOHYD 38 38 POTENTIAL.  
FT CARBOHYD 60 60 POTENTIAL.  
FT CARBOHYD 126 126 POTENTIAL.  
FT CARBOHYD 181 181 POTENTIAL.  
SQ SEQUENCE 335 AA; 37717 MW; 6B6AB99E CRC32;  
  
Query Match 2.9%; Score 7; DB 1; Length 335;  
Best Local Similarity 100.0%; Pred. No. 1.17e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 216 GPGRLLL 222  
QY 35 GPGRLLL 41  
|||||||  
RESULT 12  
ID WNSA\_AMBME STANDARD; PRT; 359 AA.  
AC Q06442;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE WNT-5A PROTEIN PRECURSOR.  
GN WNT-5A.  
OS AMBYSTOMA MEXICANUM (AXOLOTL).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA;  
OC CAUDATA; SALAMANDROIDEA; AMBYSTOMATIDAE; AMBYSTOMA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93183769.  
RA BUSSE U., SEGUIN C.;  
RT "Isolation of cDNAs for two closely related members of the axolotl  
wnt family, Awnt-5A and Awnt-5B, and analysis of their expression  
during development";  
RL MECH. DEV. 40:63-72(1993).  
CC -!- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING  
MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE REGIONS OF  
TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS.  
CC -!- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE  
EXTRACELLULAR MATRIX.  
CC -!- TISSUE SPECIFICITY: NEUROECTODERMAL AND NON-NEUROECTODERMAL  
TISSUES.  
CC -!- DEVELOPMENTAL STAGE: ABUNDANT IN THE BLASTULA UNTIL GASTRULATION,  
BARELY DETECTABLE DURING GASTRULATION, AND INCREASE AGAIN DURING  
NEURULATION. DETECTED THROUGHOUT THE REMAINING DEVELOPMENT AND IN  
HATCHED LARVAE.  
CC -!- SIMILARITY: BELONGS TO THE WNT FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; Z14047; G62427;  
DR PIR; S24999; S24999.  
DR PROSITE; PS00246; WNT1; 1.  
DR PFAM; PF00110; wnt; 1.  
KW DEVELOPMENTAL PROTEIN; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 20

```
FT CHAIN 21 359 WNT-5A PROTEIN.
FT CARBOHYD 93 93 POTENTIAL.
FT CARBOHYD 99 99 POTENTIAL.
FT CARBOHYD 291 291 POTENTIAL.
FT CARBOHYD 305 305 POTENTIAL.
SQ SEQUENCE 359 AA; 40116 MW; 7A997692 CRC32;

Query Match 2.9%; Score 7; DB 1; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.17e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 10 LALLCAL 16
| | | | |
Qy 16 LALLCAL 22

RESULT 13
ID PILC_PSEPU STANDARD; PRT; 401 AA.
AC P36641;
01-JUN-1994 (REL. 29, CREATED)
01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE FIMBRIAL ASSEMBLY PROTEIN PILC.
GN PILC.
OS PSEUDOMONAS PUTIDA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
CC PSEUDOMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WCS358;
RA MEDLINE: 94131942.
RX DE GROOT A., HELJEN I., DE COCK H., FILLOUX A., TOMMASSEN J.;
RT "Characterization of type IV pilus genes in plant growth-promoting
Pseudomonas putida WCS358."
RL J. BACTERIOL. 176:642-650(1994).
CC -!- FUNCTION: INVOLVED IN THE TRANSLOCATION OF THE TYPE IV PILIN
(PILA).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
(PROBABLY).
CC -!- SIMILARITY: BELONGS TO THE PULF/OUTF/EXEF/XPSF/XPCS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; X74276; G396264; -
PIR; S35952; S35952.
DR PROSITE; PS00874; T2SP.F; 1.
DR PFAM; PF00482; GSP11.F; 1.
KW TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE; FIMBRIA.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 375 395 POTENTIAL.
SQ SEQUENCE 401 AA; 43084 MW; A9479101 CRC32;

Query Match 2.9%; Score 7; DB 1; Length 401;
Best Local Similarity 100.0%; Pred. No. 1.17e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 229 VLLAVAA 235
| | | | |
Qy 169 VLLAVAA 175

RESULT 14
ID VGLG_IHNV STANDARD; PRT; 508 AA.
AC P07923;
01-AUG-1988 (REL. 08, CREATED)
01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DE SPIKE GLYCOPROTEIN PRECURSOR.
GN SPIKE GLYCOPROTEIN PRECURSOR.
OS INFECTIOUS HEMATOPOIETIC NECROSIS VIRUS (STRAIN ROUND BUTTE) (IHNV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
RHABDOVIRIDAE; UNCLASSIFIED RHABDOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87198856.
RA KOENER J.F., PASSAVANT C.W., KURATH G., LEONG J.;
RT "Nucleotide sequence of a cDNA clone carrying the glycoprotein gene
of infectious hematopoietic necrosis virus, a fish rhabdovirus."
RL J. VIROL. 61:1342-1349(1987).
RN [2]
RP REVISIONS.
RA LEONG J.;
RL SUBMITTED (XXX-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; M16023; G1009198; -
PIR; A29532; VGVNFR.
DR PFAM; PF00974; Rhabd_glycop; 1.
KW TRANSMEMBRANE; ENVELOPE PROTEIN; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 20
FT CHAIN 21 508 SPIKE GLYCOPROTEIN.
FT CARBOHYD 56 56 POTENTIAL.
FT CARBOHYD 400 400 POTENTIAL.
FT CARBOHYD 401 401 POTENTIAL.
FT CARBOHYD 438 438 POTENTIAL.
FT CARBOHYD 506 506 POTENTIAL.
SQ SEQUENCE 508 AA; 56799 MW; 3B4EC876 CRC32;

Query Match 2.9%; Score 7; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.17e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 477 LLAVAAC 483
| | | | |
Qy 170 LLAVAAC 176

RESULT 15
ID VGLG_CHAV STANDARD; PRT; 524 AA.
AC P13180;
01-JAN-1990 (REL. 13, CREATED)
01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE SPIKE GLYCOPROTEIN PRECURSOR.
GN SPIKE GLYCOPROTEIN PRECURSOR.
OS CHANDIPURA VIRUS (STRAIN I653514).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
RHABDOVIRIDAE; VESICULOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89299473.
RA MASTERS P.S., BHELLA R.S., BUTCHER M., PATEL B., GHOSH H.P.,
RA BANERJEE A.K.;
RT "Structure and expression of the glycoprotein gene of Chandipura
virus."
RL VIROLOGY 171:285-290(1989).
CC -!- FUNCTION: THIS PROTEIN FORMS SPIKES ON THE SURFACE OF THE VIRION.
CC IT IS RESPONSIBLE BOTH FOR THE BINDING OF THE VIRUS TO SUSCEPTIBLE
CC HOST CELLS AND FOR INDUCING THE UPTAKE OF THE VIRUS BY THE CELL.
CC THE INTERACTION BETWEEN THE INTERNAL COMPONENTS OF THE VIRION
CC AND THE PORTION OF THE GLYCOPROTEIN EXPOSED ON THE CYTOPLASMIC
CC FACE OF THE PLASMA MEMBRANE PROBABLY DIRECTS ENVELOPMENT AND
```

```
CC VIRUS BUDDING.
CC -!- SUBUNIT: TRIMERS IN THE ENDOPLASMIC RETICULUM.
CC -!- PTM: THIS PROTEIN IS MODIFIED BY THE COVALENT ADDITION OF PALMITIC
CC ACID VIA A THIOETHER LINKAGE TO A CYSTEINE. IT COULD BE EITHER ON
CC POSITION 479 OR 484.
CC -!- SIMILARITY: 39% IDENTITY TO THE G PROTEINS OF VSV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J04350; G323377; -.
CC DR PIR: A32443; VGVNCV.
CC DR PFAM: PF00974; Rhabd_glycop: 1.
CC TRN TRANSMEMBRANE; ENVELOPE PROTEIN; GLYCOPROTEIN; LIPOPROTEIN; SIGNAL.
CC SIGNAL 1 21
CC CHAIN 22 524 SPIKE GLYCOPROTEIN.
CC DOMAIN 22 472 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 473 496 POTENTIAL.
CC DOMAIN 497 524 CYTOPLASMIC (POTENTIAL).
CC CARBOHYD 184 184 POTENTIAL.
CC CARBOHYD 344 344 POTENTIAL.
CC LIPID 479 479 PALMITATE (POTENTIAL).
CC LIPID 484 484 PALMITATE (POTENTIAL).
CC SEQUENCE 524 AA; 58826 MW; AA3EB9E1 CRC32;
CC -----
Query Match 2.9%; Score 7; DB 1; Length 524;
Best Local Similarity 100.0%; Pred. No. 1.17e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 501 EERERS 507
Oy 223 EERERS 229
|||||
```

Search completed: Mon Jul 12 12:14:01 1999  
Job time : 17 secs.

\*\*\*\*\*

WQSEFH  
(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

apsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
on: Mon Jul 12 12:10:14 1999; MasPar time 16.98 Seconds  
732.927 Million cell updates/sec  
Regular output not generated.

Title: >US-08-911-423-2  
Description: (1-228) from US08911423.pep  
Perfect Score: 228  
Sequence: 1 MGAWAMLYGVSMCLVDLGO.....PEERGGQTEKCHLGRWP 228

Scoring table: TABLE unitprotatable  
Gap 60

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database: sptrenb19

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 3.380; Variance 0.386; scale 8.752

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	228	100.0	228	11	Q35714	GLUCOCORTICOID INDUCED	0.006+00	
2	7	3.1	120	2	Q53783	VALINE DEHYDROGENASE (	4.538-01	
3	7	3.1	137	14	Q72584	POLYPROTEIN (FRAGMENT)	4.538-01	
4	7	3.1	436	3	Q36020	HYPOTHETICAL 50.1 KD P	4.538-01	
5	7	3.1	453	5	Q76273	POLAR TUBE PROTEIN PTP	4.538-01	
6	7	3.1	472	2	P95423	MEKE, MEXF & OPRN GENE	4.538-01	
7	7	3.1	474	14	O55599	REPLICASE (FRAGMENT)	4.538-01	
8	7	3.1	791	2	O83389	HYPOTHETICAL 89.5 KD P	4.538-01	
9	7	3.1	1237	14	O56042	SEGMENT S1 P1.	4.538-01	
10	7	3.1	1543	14	O09498	REPLICASE.	4.538-01	
11	6	2.6	35	14	O91403	ENVELOPE GLYCOPROTEIN	4.028+01	
12	6	2.6	80	10	Q42085	ENDOPEPTIDINASE PRECURSOR	4.028+01	
13	6	2.6	92	2	O68260	YOJE.	4.028+01	
14	6	2.6	94	10	Q42339	TRANSDUCCIN HOMOLOGUE (	4.028+01	
15	6	2.6	95	11	O08843	LEUCOCYTE SPECIFIC TRA	4.028+01	
16	6	2.6	97	2	O54568	M TYPE PT4245 (EMML)	4.028+01	
17	6	2.6	141	2	Q46343	CYTOTOXIN GENE.	4.028+01	
18	6	2.6	161	2	Q48805	SBPA.	4.028+01	
19	6	2.6	161	2	P72870	ALLOPHYCOCYANIN-B.	4.028+01	
20	6	2.6	177	14	O72353	COAT PROTEIN (FRAGMENT	4.028+01	

21	6	2.6	186	5	Q18232	C26F1.3 PROTEIN.	4.02e+01
22	6	2.6	189	14	O68959	NON-STRUCTURAL PROTEIN	4.02e+01
23	6	2.6	201	3	O06252	SIMILARITY TO 23 KD MO	4.02e+01
24	6	2.6	213	10	O81431	T24H24.20 PROTEIN.	4.02e+01
25	6	2.6	225	14	Q76639	NEF PROTEIN.	4.02e+01
26	6	2.6	246	4	Q14865	MODULATOR RECOGNITION	4.02e+01
27	6	2.6	253	1	O28290	PEPTIDYL-PROLYL CIS-TR	4.02e+01
28	6	2.6	259	2	O54168	TRYPSINOGEN PRECURSOR.	4.02e+01
29	6	2.6	274	10	O23248	CLASS IV CHITINASE (EC	4.02e+01
30	6	2.6	287	10	Q38818	CYCLIN 2.	4.02e+01
31	6	2.6	308	2	O50418	NUCLEOSIDE HYDROLASE.	4.02e+01
32	6	2.6	312	2	O54437	MALONYL-COA:ACYL CARRI	4.02e+01
33	6	2.6	313	2	O85702	PUTATIVE OXIDOREDUCTAS	4.02e+01
34	6	2.6	338	5	Q17167	ORF1 (BMC7).	4.02e+01
35	6	2.6	343	5	O18070	T08G3.2 PROTEIN.	4.02e+01
36	6	2.6	343	14	Q04044	NONSTRUCTURAL PROTEIN	4.02e+01
37	6	2.6	343	14	O81747	NONSTRUCTURAL PROTEIN	4.02e+01
38	6	2.6	359	1	O27738	HYPOTHETICAL 39.7 KD P	4.02e+01
39	6	2.6	368	13	Q90853	HOMEOBOX PROTEIN	4.02e+01
40	6	2.6	387	3	Q13918	PROBABLE ZINC/CADMIUM	4.02e+01
41	6	2.6	407	4	O00536	TTF-1 INTERACTING PEPT	4.02e+01
42	6	2.6	414	1	O58491	414AA LONG HYPOTHETICA	4.02e+01
43	6	2.6	427	14	O68483	POLYPROTEIN (FRAGMENT)	4.02e+01
44	6	2.6	427	14	O68492	POLYPROTEIN (FRAGMENT)	4.02e+01
45	6	2.6	427	14	O68456	POLYPROTEIN (FRAGMENT)	4.02e+01
46	6	2.6	427	14	O68484	POLYPROTEIN (FRAGMENT)	4.02e+01
47	6	2.6	427	14	O68485	POLYPROTEIN (FRAGMENT)	4.02e+01
48	6	2.6	427	14	O68490	POLYPROTEIN (FRAGMENT)	4.02e+01
49	6	2.6	427	14	O68493	POLYPROTEIN (FRAGMENT)	4.02e+01
50	6	2.6	427	14	O68487	POLYPROTEIN (FRAGMENT)	4.02e+01
51	6	2.6	427	14	O68464	POLYPROTEIN (FRAGMENT)	4.02e+01
52	6	2.6	427	14	O68486	POLYPROTEIN (FRAGMENT)	4.02e+01
53	6	2.6	437	2	O07462	BENZYL-COA REDUCTASE	4.02e+01
54	6	2.6	465	2	P74491	HYPOTHETICAL 50.3 KD P	4.02e+01
55	6	2.6	489	2	P74573	LIM17	4.02e+01
56	6	2.6	506	11	O88562	FATTY ACID TRANSPORT P	4.02e+01
57	6	2.6	535	5	O16778	ROT3.1 PROTEIN.	4.02e+01
58	6	2.6	571	10	Q42608	PECTINESTERASE (EC 3.1	4.02e+01
59	6	2.6	660	10	O23435	HYPOTHETICAL PROTEIN.	4.02e+01
60	6	2.6	688	14	O85646	MAMMARY TUMOR VIRUS LO	4.02e+01
61	6	2.6	706	5	Q18455	C34E11.3 (FRAGMENT).	4.02e+01
62	6	2.6	746	14	O67813	VIRAL POLYPROTEIN (FRA	4.02e+01
63	6	2.6	829	4	O75309	KSP-CADHERIN.	4.02e+01
64	6	2.6	837	4	O60570	ADAPTOR PROTEIN X11ALP	4.02e+01
65	6	2.6	839	11	Q35430	NEURONAL MUNC18-1 BIND	4.02e+01
66	6	2.6	858	14	O69535	STRUCTURAL PROTEIN.	4.02e+01
67	6	2.6	899	11	Q63527	RETINOBLASTOMA PROTEIN	4.02e+01
68	6	2.6	951	2	O68078	ICE NUCLEATION PROTEIN	4.02e+01
69	6	2.6	961	10	O22937	T11A07.7 PROTEIN.	4.02e+01
70	6	2.6	989	2	O59736	2-OXOGlutARATE DEHYDRO	4.02e+01
71	6	2.6	1025	10	P93067	CALMODULIN-STIMULATED	4.02e+01
72	6	2.6	1201	5	Q24240	ACTIN BINDING PROTEIN.	4.02e+01
73	6	2.6	1355	13	Q42350	ALPHA 2 TYPE I COLLAGE	4.02e+01
74	6	2.6	1483	5	O22528	T16G12.5 PROTEIN.	4.02e+01
75	6	2.6	1805	14	Q41809	POLYPROTEIN.	4.02e+01
76	6	2.6	2109	13	P79787	CHONDROITIN SULFATE PR	4.02e+01
77	6	2.6	2218	14	O67817	COMPLETE GENOME.	4.02e+01
78	6	2.6	2218	14	O67824	GBM/FRHK RNA.	4.02e+01
79	6	2.6	2284	14	O81817	POLYPROTEIN PRECURSOR.	4.02e+01
80	6	2.6	2436	14	O81756	POLYPROTEIN (FRAGMENT)	4.02e+01
81	6	2.6	3010	14	P90192	POLYPROTEIN.	4.02e+01
82	6	2.6	3010	14	P90191	POLYPROTEIN.	4.02e+01
83	6	2.6	3010	14	P90193	POLYPROTEIN.	4.02e+01
84	6	2.6	3010	14	P90195	POLYPROTEIN.	4.02e+01
85	6	2.6	3010	14	O81760	POLYPROTEIN PRECURSOR.	4.02e+01
86	6	2.6	3010	14	O68949	CORE, E1, NS1/E2, NS2,	4.02e+01
87	6	2.6	3010	14	O93016	POLYPROTEIN.	4.02e+01
88	6	2.6	3010	14	O81541	POLYPROTEIN.	4.02e+01
89	6	2.6	3010	14	P88803	RNA FOR POLYPROTEIN, C	4.02e+01
90	6	2.6	3010	14	P89966	POLYPROTEIN.	4.02e+01
91	6	2.6	3010	14	O68826	HCV POLYPROTEIN.	4.02e+01
92	6	2.6	3010	14	O68788	POLYPROTEIN.	4.02e+01
93	6	2.6	3010	14	O81757	POLYPROTEIN.	4.02e+01

94 6 2.6 3010 14 P90194 POLYPROTEIN. 4.02e+01  
95 6 2.6 3011 14 Q81754 POLYPROTEIN. 4.02e+01  
96 6 2.6 3011 14 Q36610 POLYPROTEIN. 4.02e+01  
97 6 2.6 3011 14 Q36608 POLYPROTEIN. 4.02e+01  
98 6 2.6 3011 14 Q36609 POLYPROTEIN. 4.02e+01  
99 6 2.6 4063 5 O61612 BETA CHAIN SPECTRIN HO 4.02e+01  
100 6 2.6 4151 2 O53490 POLYKETIDE SYNTHASE. 4.02e+01

## ALIGNMENTS

RESULT 1  
ID O35714 PRELIMINARY; PRT; 228 AA.  
AC O35714;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE GLUCOCORTICOID INDUCED TNF FAMILY RELATED PROTEIN PRECURSOR.  
GN GTR.  
MUS MUSCULUS (MOUSE).  
EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN=C3H/HEN;  
RX MEDLINE: 97322352.  
RA NOCENTINI G., GIUNCHI L., RONCHETTI S., KRAUSZ L.T., BARTOLI A.,  
RA MORACA R., MIGLIORATI G., RICCARDI C.;  
RT "A new member of the tumor necrosis factor/nerve growth factor  
RT receptor family inhibits T cell receptor-induced apoptosis."  
RL PROC. NATL. ACAD. SCI. U.S.A. 94:6216-6221(1997).  
DR EMBL: U82534; G2228584;  
DR MGD: MGI:894675; GTR.  
KW SIGNAL.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 228 GLUCOCORTICOID INDUCED TNF FAMILY  
FT RELATED PROTEIN.  
SQ SEQUENCE 228 AA; 25334 MW; BA433757 CRC32;

Query Match 100.0%; Score 228; DB 11; Length 228;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MGAWMLYGVSMCLVDLGQPSVVEPCGCGKQVNGSGNNTCCSLYAPGKEDCPKRC 60  
Qy 1 MGAWMLYGVSMCLVDLGQPSVVEPCGCGKQVNGSGNNTCCSLYAPGKEDCPKRC 60  
61 ICVTPEYHCGDPQCKICKHYPCQQRVESQGDIVFGFRCVACAMGTFSGRDGHCLWT 120  
61 ICVTPEYHCGDPQCKICKHYPCQQRVESQGDIVFGFRCVACAMGTFSGRDGHCLWT 120  
121 NCSQFGFLTFMPGNKTHNAVCIPELPTQYQHLTVIFLVMAACIFFLTVQLGHLHWL 180  
121 NCSQFGFLTFMPGNKTHNAVCIPELPTQYQHLTVIFLVMAACIFFLTVQLGHLHWL 180  
181 RQHMCPRETQFFAEVQLSAEDACSFQPEERGGQTEKCHLGRWP 228  
181 RQHMCPRETQFFAEVQLSAEDACSFQPEERGGQTEKCHLGRWP 228

RESULT 2  
ID Q53783 PRELIMINARY; PRT; 120 AA.  
AC Q53783;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
DE VALINE DEHYDROGENASE (VDH) (FRAGMENT).  
OS STREPTOMYCES AMBOFACIENS.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE: 95014047.

RA TANG L., ZHANG Y.X., HUTCHINSON C.R.;  
RT "Amino acid catabolism and antibiotic synthesis: valine is a source of  
RT precursors for macrolide biosynthesis in Streptomyces ambofaciens and  
RT Streptomyces fradiae."  
RL J. BACTERIOL. 176:6107-6119(1994).  
DR EMBL: L33871; G688448;  
FT NON\_TER 120 120  
SQ SEQUENCE 120 AA; 13690 MW; AEA28DBB CRC32;

Query Match 3.1%; Score 7; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 4.53e-01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 24 FPEERG 30  
Qy 208 FPEERG 214

RESULT 3  
ID O72584 PRELIMINARY; PRT; 137 AA.  
AC O72584;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE POLYPROTEIN (FRAGMENT).  
OS SINDBIS VIRUS.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;  
OC ALPHAVIRUS.  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN=K10390, AUSTRALIAN ISOLATE;  
RA SAMMELS L.M., LINDSAY M.D., POIDINGER M., COELEN R.J.,  
RA MACKENZIE J.S.;  
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF061712; G3136256;  
KW POLYPROTEIN.  
FT NON\_TER 1 1  
FT NON\_TER 137 137  
SQ SEQUENCE 137 AA; 15216 MW; OE922820 CRC32;

Query Match 3.1%; Score 7; DB 14; Length 137;  
Best Local Similarity 100.0%; Pred. No. 4.53e-01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 130 FSAGRDG 136  
Qy 108 FSAGRDG 114

RESULT 4  
ID O36020 PRELIMINARY; PRT; 436 AA.  
AC O36020;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 50.1 KD PROTEIN C4F10.08 IN CHROMOSOME I.  
GN SPAC4F10.08.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;  
OC SCHIZOSACCHAROMYCES.  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;  
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
DR EMBL: Z98980; E339955;  
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
FT TRANSMEM 269 289 POTENTIAL.  
FT TRANSMEM 305 325 POTENTIAL.  
FT TRANSMEM 341 361 POTENTIAL.  
FT TRANSMEM 371 391 POTENTIAL.

SQ SEQUENCE 436 AA; 50113 MW; DDD9A105 CRC32;

Query Match 3.1%; Score 7; DB 3; Length 436;  
Best Local Similarity 100.0%; Pred. No. 4.53e-01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 348 ACIFILT 354

QY 163 ACIFILT 169

RESULT 5

ID 076273 PRELIMINARY; PRT; 453 AA.

AC 076273;

DT 01-NOV-1998 (TREMBLREL. 08, CREATED)

DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE POLAR TUBE PROTEIN PTP55 PRECURSOR.

EN PTP55.

ENCEPHALITOZOON HELLEM.

EUKARYOTA; MICROSPORIDIA; UNIKARYONIDAE; ENCEPHALITOZOON.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 98418769.

RA KEHANE E.M., ORR G.A., ZHANG H.S., TAKVORIAN P.M., CALI A.,

RA TANOWITZ H.B., WITTNER M., WEISS L.M.;

RT "The molecular characterization of the major polar tube protein gene

from Encephalitozoon hellem, a microsporidian parasite of humans."

RL MOL. BIOCHEM. PARASITOL. 94:227-236(1998).

DR EMBL; AF044915; G3452285; "

KW SIGNAL.

FT SIGNAL. 1 22 POTENTIAL.

FT CHAIN 23 453 POLAR TUBE PROTEIN PTP55.

SQ SEQUENCE 453 AA; 45461 MW; BIC80D44 CRC32;

Query Match

Best Local Similarity 3.1%; Score 7; DB 5; Length 453;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 233 YPCPGQ 239

QY 80 YPCPGQ 86

RESULT 6

ID P95423 PRELIMINARY; PRT; 472 AA.

AC P95423;

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE MEXE, MEXF & OPRN GENES.

GN OPRN.

OS PSEUDOMONAS AERUGINOSA.

OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;

CC PSEUDOMONAS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PA01;

RX MEDLINE; 97197179.

RA KOHLER T., MICHEA-HAMZEPPOUR M., HENZE U., GOTOH N., CURTY L.K.,

RA PECHERE J.C.;

RT "Characterization of MexE-MexF-OprN, a positively regulated multidrug

efflux system of Pseudomonas aeruginosa."

RL MOL. MICROBIOL. 23:345-354(1997).

DR EMBL; X9514; E256816; "

SQ SEQUENCE 472 AA; 51338 MW; E1C9E8B8 CRC32;

Query Match

Best Local Similarity 3.1%; Score 7; DB 2; Length 472;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 439 QLSAEDA 445

QY 439 QLSAEDA 445

QY 197 QLSAEDA 203

RESULT 7

ID 055599 PRELIMINARY; PRT; 474 AA.

AC 055599;

DT 01-JUN-1998 (TREMBLREL. 06, CREATED)

DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE REPLICASE (FRAGMENT).

OS GARLIC VIRUS B.

OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE.

RN [1]

RP SEQUENCE FROM N.A.

RA SUMI S.;

RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 93389442.

RA SUMI S., TSUNEYOSHI T., FURUTANI H.;

RT "Novel rod-shaped viruses isolated from garlic, Allium sativum,

possessing a unique genome organization."

RL J. GEN. VIROL. 74:1879-1885(1993).

DR EMBL; AB010301; D1025473; "

FT NON-TER 1

SQ SEQUENCE 474 AA; 54242 MW; 1D4AFB90 CRC32;

Query Match

Best Local Similarity 3.1%; Score 7; DB 14; Length 474;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 54 PRETQPF 60

QY 187 PRETQPF 193

RESULT 8

ID 083389 PRELIMINARY; PRT; 791 AA.

AC 083389;

DT 01-NOV-1998 (TREMBLREL. 08, CREATED)

DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 89.5 KD PROTEIN.

GN TP0374.

OS TREPONEMA PALLIDUM.

OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 98332770.

RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,

RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,

RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,

RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,

RA McDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,

RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,

RA VENTER J.C.;

RT "Complete Genome Sequence of Treponema pallidum, the Syphilis

Spirochete."

RL SCIENCE 281:375-388(1998).

RN [2]

RP SEQUENCE FROM N.A.

RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,

RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,

RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,

RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,

RA McDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,

RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,

RA VENTER J.C.;

RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AE001216; G3322655; "

KW HYPOTHETICAL PROTEIN.

SQ SEQUENCE 791 AA; 89451 MW; E3E1DE4C CRC32;

Query Match 3.1%; Score 7; DB 2; Length 791;  
Best Local Similarity 100.0%; Pred. No. 4.53e-01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 559 EPLPTEQ 565  
QY 144 EPLPTEQ 150

RESULT 9  
ID O56042 PRELIMINARY; PRT; 1237 AA.  
AC O56042;  
DT 01-JUN-1998 (TREMREL. 06, CREATED)  
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
DE SEGMENT S1 P1.  
OS RICE RAGGED STUNT VIRUS.  
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ORYZAVIRUS.  
RN [1]  
SEQUENCE FROM N.A.  
UPADHYAYA N.M., LI Z., RAMM K., YANG M., GELLATLY J.A.,  
KOSTRATANA W., GERLACH W.L., WATERHOUSE P.M.;  
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF020334; G2921096; -.  
SQ SEQUENCE 1237 AA; 137684 MW; 7933CB71 CRC32;

Query Match 3.1%; Score 7; DB 14; Length 1237;  
Best Local Similarity 100.0%; Pred. No. 4.53e-01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 283 QLSAEDA 289  
QY 197 QLSAEDA 203

RESULT 10  
ID O09498 PRELIMINARY; PRT; 1543 AA.  
AC O09498;  
DT 01-JUL-1997 (TREMREL. 04, CREATED)  
DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE REPLICASE.  
OS GARLIC VIRUS B.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE.  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN-KOREA;  
MEDLINE; 98120835.  
SONG S.I., SONG J.T., KIM C.H., LEE J.S., CHOI Y.D.;  
"Molecular characterization of the garlic virus X genome.";  
J. GEN. VIROL. 79:155-159(1998).  
RN [2]  
SEQUENCE FROM N.A.  
RC STRAIN-KOREA;  
RA CHOI Y.D.;  
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U89243; G1888229; -.  
SQ SEQUENCE 1543 AA; 174307 MW; 0F7F14BC CRC32;

Query Match 3.1%; Score 7; DB 14; Length 1543;  
Best Local Similarity 100.0%; Pred. No. 4.53e-01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1124 PRETOPF 1130  
QY 187 PRETOPF 193

RESULT 11  
ID O91403 PRELIMINARY; PRT; 35 AA.  
AC O91403;  
DT 01-NOV-1998 (TREMREL. 08, CREATED)  
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).  
RN [1]  
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).  
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TK-29;  
RX MEDLINE; 98090117.  
RA IDA S., GATANAGA H., SHIODA T., NAGAI Y., KOBAYASHI N., SHIMADA K.,  
RA KIMURA S., IWAMOTO A., OKA S.;  
"HIV type 1 V3 variation dynamics in vivo: long-term persistence of  
non-synctium-inducing genotypes and transient presence of  
ret-cyctium-inducing genotypes during the course of progressive AIDS.";  
RL AIDS RES. HUM. RETROVIRUSES 13:1597-1609(1997).  
DR EMBL; AB005375; D1034259; -.  
KW ENVELOPE PROTEIN.  
FT NON\_TER 1 1  
FT NON\_TER 35 35  
SQ SEQUENCE 35 AA; 3817 MW; 23239C4B CRC32;

Query Match 2.6%; Score 6; DB 14; Length 35;  
Best Local Similarity 100.0%; Pred. No. 4.02e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 4 SGNNT 9  
QY 38 SGNNT 43

RESULT 12  
ID Q42085 PRELIMINARY; PRT; 80 AA.  
AC Q42085;  
DT 01-NOV-1996 (TREMREL. 01, CREATED)  
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)  
DE ENDOCHITINASE PRECURSOR (FRAGMENT).  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
RN [1]  
SEQUENCE FROM N.A.  
RP TISSUE-CELL SUSPENSION CULTURE OF A. THALIANA ECOTYPE;  
RA PHILIPPS G., GIGOT C.;  
RL SUBMITTED (SEP-1993) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; Z26409; G403221; -.  
DR PROSITE; PS00026; CHITIN BINDING; 1.  
DR PFAM; PF00187; chitin binding; 1.  
DR MENDEL; 14795; ARATH; Chia0; mm14795.  
KW CHITIN-BINDING.  
FT NON\_TER 1 1  
FT NON\_TER 80 80  
SQ SEQUENCE 80 AA; 8326 MW; E6968EBF CRC32;

Query Match 2.6%; Score 6; DB 10; Length 80;  
Best Local Similarity 100.0%; Pred. No. 4.02e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 47 CSQFGF 52  
QY 122 CSQFGF 127

RESULT 13  
ID O68260 PRELIMINARY; PRT; 92 AA.  
AC O68260;  
DT 01-AUG-1998 (TREMREL. 07, CREATED)  
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
DE YOJE.  
GN YOJE.  
OS BACILLUS SUBTILIS.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
OC BACILLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA PARK S.-H., SHIN B.-S., CHOI S.-K., GHIM S.-Y.;  
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF026147; G3169321; -.  
SQ SEQUENCE 92 AA; 10750 MW; 39CA796B CRC32;

Query Match 2.6%; Score 6; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 4.02e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 86 EERGEQ 91  
QY 211 EERGEQ 216  
|||||

RESULT 14  
Q42339 PRELIMINARY; PRT; 94 AA.  
AC Q42339;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
DE TRANSDUCIN HOMOLOGUE (FRAGMENT).  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-CELL SUSPENSION CULTURE OF ECOTYPE COLUMBIA;  
RA COOKE R., LAUDIE M., RAYNAL M., DELSENY M.;  
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; F19990; E225583; -.  
DR MENDEL; 13102; ARATH;2302;1.  
FT NON\_TER 1  
NON\_TER 94 94  
SQ SEQUENCE 94 AA; 10185 MW; 18AF2A28 CRC32;

Query Match 2.6%; Score 6; DB 10; Length 94;  
Best Local Similarity 100.0%; Pred. No. 4.02e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 FSAGRD 31  
QY 108 FSAGRD 113  
|||||

RESULT 15  
ID C08843 PRELIMINARY; PRT; 95 AA.  
AC C08843;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE LEUCOCYTE SPECIFIC TRANSCRIPT 1 (B144).  
GN LST1 OR B144.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98035883.  
RA DE BAIEY A., FELLERHOFF B., MAIER S., MARTINOZZI S., WEIDLE U.,  
RA WEISS E.H.;  
RT "Complex expression pattern of the TNF region gene LST1 through  
RT differential regulation, initiation, and alternative splicing.";  
RL GENOMICS 45:591-600(1997).  
DR EMBL; AF000427; G2145070; -.  
DR MGD; MGI:1096324; LST1.  
SQ SEQUENCE 95 AA; 10325 MW; 9879ED21 CRC32;

Query Match 2.6%; Score 6; DB 11; Length 95;  
Best Local Similarity 100.0%; Pred. No. 4.02e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 NGSNN 10  
QY 36 NGSNN 41  
|||||

Search completed: Mon Jul 12 12:10:39 1999  
Job time : 25 secs.

\*\*\*\*\*

MAQSGRAH  
(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
on: Mon Jul 12 12:12:17 1999; MasPar time 15.87 Seconds  
Tabular output not generated. 323.021 Million cell updates/sec

Title: >US-08-911-423-4  
Description: (1-241) from US08911423.pap  
Perfect Score: 241  
Sequence: 1 MAQHGAGAFALCGLALLC.....EEEGERSAEEKRGLGLDW 241

Scoring table: TABLE unitprotable  
Gap 60

Searched: 170751 seqs, 21266608 residues

Post-Processing: Minimum Match 0%  
Listing first 100 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 2.772; Variance 0.718; scale 3.860

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	241	100.0	241	31	W37839 Amino acid sequence o	0.00e+00
2	194	80.5	311	31	W37842 Human 312C2 protein f	4.51e-299
3	187	77.6	228	31	W37840 Truncated human 312C2	1.71e-287
4	99	41.1	232	31	W37841 Human 312C2 protein f	1.20e-142
5	11	4.6	222	33	W49018 Mouse glucocorticoid	2.74e-04
6	11	4.6	228	33	W49016 Mouse glucocorticoid	2.74e-04
7	11	4.6	228	31	W37838 Amino acid sequence o	2.74e-04
8	11	4.6	294	33	W49017 Mouse glucocorticoid	2.74e-04
9	7	2.9	79	36	W70320 Secreted protein B011	2.05e-01
10	7	2.9	138	5	R26820 CA455 protein.	2.05e-01
11	7	2.9	428	33	W63713 Human hki1 protein.	2.05e-01
12	7	2.9	498	31	W53461 Human latheo protein.	2.05e-01
13	7	2.9	556	31	W53459 Human latheo protein	2.05e-01
14	7	2.9	713	23	W25638 Human cadherin-13.	2.05e-01
15	7	2.9	713	21	W13336 Full length human cad	2.05e-01
16	7	2.9	1208	28	W40827 Human Jagged protein.	2.05e-01

1218	19	W05833	Human Serrate-1 (HJ1)	2.05e+01
1218	25	W18354	Proliferation and dif	2.05e+01
1218	29	W44301	Human serrate 1.	2.37e+02
13	5	R26992	IGFBP-6 amino termina	2.37e+02
14	17	R98180	Peptide fragment of t	2.37e+02
15	9	R47690	HIV epitope #23.	2.37e+02
20	2	P70701	Sequence encoded by t	2.37e+02
24	27	W45016	Immunomodulatory pep	2.37e+02
25	6	R30690	Synthetic LamB4 signa	2.37e+02
26	6	R86600	gp41 Epitope-substitu	2.37e+02
35	17	R88632	Secreted protein enco	2.37e+02
38	38	W59432	Human chemokine prote	2.37e+02
86	32	W59432	Human macrophage deri	2.37e+02
93	23	W20058	Macrophage derived ch	2.37e+02
93	23	W20058	Hepatitis GB virus (H	2.37e+02
101	16	R81422	A cysteine rich solub	2.37e+02
108	39	W87710	High-affinity branch	2.37e+02
160	34	W38571	Protein sequence enco	2.37e+02
163	29	W46880	Wild tomato acyltras	2.37e+02
167	37	W80477	Wild tomato acyltras	2.37e+02
179	29	W41578	tICAM(185) with an ad	2.37e+02
186	27	W42407	Sequence encoded by t	2.37e+02
213	2	P70700	32K S2 protein encode	2.37e+02
214	2	P70154	pMS18 contg. male fio	2.37e+02
221	26	R06453	Sequence encoded by n	2.37e+02
222	2	P70699	Rabbit G-protein coup	2.37e+02
252	17	R91232	Hepatitis GB virus (H	2.37e+02
265	16	R81426	Babesia equi merozoit	2.37e+02
277	23	W18875	Membrane-type matrix	2.37e+02
324	37	W79300	Trichoderma reesei AC	2.37e+02
341	33	W58573	WD-40 domain-contg. S	2.37e+02
341	17	R85886	Arabidopsis SMR enzym	2.37e+02
361	39	W70837	Human retroviral rece	2.37e+02
367	38	W67472	Sequence of rennin en	2.37e+02
381	4	P40218	B. steatothermophilus	2.37e+02
400	37	W68552	Human T-cell transmem	2.37e+02
453	38	W67475	E38C mutant of trunc	2.37e+02
453	10	R53467	Pseudomonas aerugin	2.37e+02
461	25	W19276	Triticum sp. cysteine	2.37e+02
469	39	W89560	G380C mutant of inter	2.37e+02
507	27	W42332	W382C mutant of inter	2.37e+02
507	27	W42333	ICAM-1.	2.37e+02
507	9	R48038	K306C mutant of inter	2.37e+02
507	27	W42326	T429C mutant of inter	2.37e+02
507	27	W42334	Human intercellular a	2.37e+02
508	35	W71269	Canine ICAM-R polypep	2.37e+02
528	38	W81457	Intracellular adhesio	2.37e+02
532	38	W70871	ICAM-1.	2.37e+02
532	16	R90294	Human intracellular a	2.37e+02
532	14	R79457	Human ICAM-1.	2.37e+02
532	27	W27270	Human ICAM-1.	2.37e+02
532	21	W14720	Calif IAP.	2.37e+02
533	8	R41666	Human intercellular a	2.37e+02
547	38	W81440	Human ICAM-R protein.	2.37e+02
547	35	W75118	Herbicide sensitive,	2.37e+02
562	1	P81153	Murine ecotropic retr	2.37e+02
622	38	W67474	Alpha4 subunit of nor	2.37e+02
627	24	W11824	Neuronal nicotinic ac	2.37e+02
627	20	W09023	Alpha4 subunit of mut	2.37e+02
627	24	W11825	Human H13 viral recep	2.37e+02
628	9	W54872	Human retroviral rece	2.37e+02
629	38	W67473	Human H13 viral recep	2.37e+02
629	9	R54870	Human H13 viral recep	2.37e+02
629	9	R54871	Antigen from cluster	2.37e+02
634	38	W89888	VCAM-6D/ICAM-2.	2.37e+02
644	13	R39685	Human sVEGF-R11.	2.37e+02
664	8	R62486	Human SRCR protein fr	2.37e+02
666	34	W64590	Soluble truncated VEG	2.37e+02
668	28	W47037	Partial thermophilic	2.37e+02
676	35	W80305	Thermotable T. aquat	2.37e+02
680	3	R15299	tICAM(453)IgG immuno	2.37e+02
680	9	R48037	Thermotoga maritima M	2.37e+02
721	36	W49862		

90 6 2.5 764 35 W69679 KDR protein sequence. 2.37e+02  
91 6 2.5 920 35 W75909 Helicobacter isoleucy 2.37e+02  
92 6 2.5 1005 32 W60181 B. thuringiensis tran 2.37e+02  
93 6 2.5 1005 12 R75354 Transposon Tn5401 tra 2.37e+02  
94 6 2.5 1356 5 R26999 Novel type III RTR en 2.37e+02  
95 6 2.5 1776 35 W50894 Mouse laminin B1 chal 2.37e+02  
96 6 2.5 1785 34 W64591 Human SRCR protein. 2.37e+02  
97 6 2.5 1958 12 R60620 Protein from ORF2 of 2.37e+02  
98 6 2.5 2251 14 R71009 Human neuronal calciu 2.37e+02  
99 6 2.5 2270 14 R71010 Human neuronal calciu 2.37e+02  
100 6 2.5 5072 2 R11510 Ryanodine receptor de 2.37e+02

## ALIGNMENTS

RESULT 1  
ID W37839 standard; Protein; 241 AA.  
AC W37839;  
DT 28-JUL-1998 (first entry)  
DE Human 312C2 T cell protein; thymus cell; spleen cell; T cell;  
KW antigen-specific T cell proliferation; cytokine production by T-cell;  
KW apoptosis; cancer; haematopoietic cells; lymphoid cell;  
KW autoimmune disorders.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT 1..726  
FT CDS /\*tag= a  
/\*product= "human 312C2 protein"  
FN W09806842-A1.  
PD 19-FEB-1998.  
PF 14-AUG-1997; U13931.  
PR 07-OCT-1996; US-027901.  
PR 16-AUG-1996; US-689943.  
FA (SCHE ) SCHERING CORP.  
PI Gorman DM, Randall TD, Zlotnik A;  
DR WPI; 98-159534/14.  
DR N-PSDB; V19153.  
PT Isolated 312C2 T cell gene - used to develop products for treating,  
PT e.g. cancers, auto-immune disorders, transplantation rejection and  
PT other T cell disorders  
PS Claim 2; Pages 59-60; 71pp; English.  
CC This is the amino acid sequence encoding the human 312C2 T cell  
CC protein. The 312C2 proteins are expressed in thymus cells and are  
CC induced on T cells and spleen cells following activation. Engagement  
CC of 312C2 stimulates proliferation of T cell clones, and potentiates T  
CC proliferation and cytokine production by T-cells, and potentiates T  
CC cell expansion or apoptosis. The products can be used in the  
CC treatment of conditions associated with abnormal physiology or  
CC development, including abnormal proliferation, e.g. cancerous  
CC conditions or degenerative conditions. They can be used in the  
CC regulation or development of haematopoietic cells, e.g. lymphoid cells  
CC which affect immunological responses, e.g. autoimmune disorders.  
SQ Sequence 241 AA;

Query Match 100.0%; Score 241; DB 31; Length 241;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 maqhamafraalcalalslgrptgpgcgrlllgtgttdarccrvhttrccrd 60  
QY 1 MAQHGMAGAFRALCGLALLCALSIGQRTGGPGCGGRLLLGTGTDARCCRVHTTRCCRD 60  
Db 61 ypgeccscwdcmvqpfhcgdpccctcrhpcppgvgvsgqkfsfgfcidcasgtf 120  
QY 61 YPGECCSEWDCMCVQPFHCGDPCCCTTCRHHPCPPGOGVQSGKFSFGFCIDCASGTF 120  
Db 121 sgghgckpwtcdtqfqltvfpnknthnvcvpgspapaelglwltvllavaacvlll 180  
QY 121 SGGHGHCKPMTDCTQFGFLTVFPNGKTHNVCVPGSPAPAEPLGLWLTVLLAVAACVLLL 180  
Db 181 tsaqllghlwlrsqcmwprretqllleppstedarscqfpeergersaeekgrldlw 240  
QY 181 TSAQLGLHIWLRSQCMWPRRETQLLLEVPSTEDARSCQFPEERGERSAEEKRLGLDW 240

QY 181 TSAQLGLHIWLRSQCMWPRRETQLLLEVPSTEDARSCQFPEERGERSAEEKRLGLDW 240  
Db 241 v 241  
QY 241 v 241  
RESULT 2  
ID W37842 standard; Protein; 311 AA.  
AC W37842;  
DT 28-JUL-1998 (first entry)  
DE Human 312C2 protein from clone.G10 amino acid sequence.  
KW Human 312C2 T cell protein; thymus cell; spleen cell; T cell;  
KW antigen-specific T cell proliferation; cytokine production by T-cell;  
KW apoptosis; cancer; haematopoietic cells; lymphoid cell;  
KW autoimmune disorders.  
OS Homo sapiens.  
FN W09806842-A1.  
PD 19-FEB-1998.  
PF 14-AUG-1997; U13931.  
PR 07-OCT-1996; US-027901.  
PR 16-AUG-1996; US-689943.  
PA (SCHE ) SCHERING CORP.  
PI Gorman DM, Randall TD, Zlotnik A;  
DR WPI; 98-159534/14.  
PT Isolated 312C2 T cell gene - used to develop products for treating,  
PT e.g. cancers, auto-immune disorders, transplantation rejection and  
PT other T cell disorders  
PS Disclosure; Pages 63-64; 71pp; English.  
CC This is the amino acid sequence of the human 312C2 T cell protein from  
CC clone.G10, which is identical to sequence W37838 for the first 202  
CC amino acid residues. The 312C2 proteins are expressed in thymus  
CC cells and are induced on T cells and spleen cells following activation.  
CC Engagement of 312C2 stimulates proliferation of T cell clones,  
CC antigen-specific proliferation and cytokine production by T-cells, and  
CC potentiates T cell expansion or apoptosis. The products can be used  
CC in the treatment of conditions associated with abnormal physiology or  
CC development, including abnormal proliferation, e.g. cancerous  
CC conditions or degenerative conditions. They can be used in the  
CC regulation or development of haematopoietic cells, e.g. lymphoid cells  
CC which affect immunological responses, e.g. autoimmune disorders.  
SQ Sequence 311 AA;

Query Match 80.5%; Score 194; DB 31; Length 311;  
Best Local Similarity 100.0%; Pred. No. 4.51e-299;  
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 maqhamafraalcalalslgrptgpgcgrlllgtgttdarccrvhttrccrdypgeec 60  
QY 7 MAQFALCGLALLCALSIGQRTGGPGCGGRLLLGTGTDARCCRVHTTRCCRDYPGEEC 66  
Db 61 csewdcmvqpfhcgdpccctcrhpcppgvgvsgqkfsfgfcidcasgtfsgghg 120  
QY 67 CSEWDCMCVQPFHCGDPCCCTTCRHHPCPPGOGVQSGKFSFGFCIDCASGTFSGGHEG 126  
Db 121 hckpwtcdtqfqltvfpnknthnvcvpgspapaelglwltvllavaacvllltsaqll 180  
QY 127 HCKPMTDCTQFGFLTVFPNGKTHNVCVPGSPAPAEPLGLWLTVLLAVAACVLLLTSQQLG 186  
Db 181 lhwqlrsqcmwpr 194  
QY 187 LHIWQLRSQCMWPR 200

RESULT 3  
ID W37840 standard; Protein; 228 AA.  
AC W37840;  
DT 28-JUL-1998 (first entry)  
DE Truncated human 312C2 protein from clone.A8 amino acid sequence.  
KW Human 312C2 T cell protein; thymus cell; spleen cell; T cell;  
KW antigen-specific T cell proliferation; cytokine production by T-cell;  
KW apoptosis; cancer; haematopoietic cells; lymphoid cell;  
KW autoimmune disorders.



KW G1TR; tumour necrosis factor; apoptosis; hodgkin's disease; G1TR-B;  
 OS G1TR-C.  
 Mus sp.  
 Key Location/Qualifiers  
 Peptide 1..19  
 Region 29..60  
 Modified\_site 36 /note= "Cysteine pseudorepeat"  
 Modified\_site 40 /note= "N-glycosylated"  
 Region 62..100 /note= "N-glycosylated"  
 Region 103..141 /note= "Cysteine pseudorepeat"  
 Modified\_site 121 /note= "Cysteine pseudorepeat"  
 Modified\_site 134 /note= "N-glycosylated"  
 Region 154..176 /note= "N-glycosylated"  
 Modified\_site 199 /note= "Probable transmembrane domain"  
 Modified\_site 199 /note= "Possibly phosphorylated"  
 WO9824895-A1.  
 11-JUN-1998.  
 08-NOV-1997; E06252.  
 02-DEC-1996; GB-025074.  
 (PHAA ) PHARMACIA & UPJOHN SPA.  
 Ricciardi C;  
 WPI; 98-333315/29.  
 N-PSDB; V32773.  
 New isolated glucocorticoid induced TNFR related polypeptide - used  
 to stimulate lymphocyte activity and cell death rescue, useful to,  
 e.g. develop products to suppress lymphocyte activity and induce  
 apoptosis  
 Claim 13: Pages 36-37; 53pp; English.  
 The present claimed sequence represents a mouse glucocorticoid  
 induced TNFR-family related protein (G1TR). The invention also  
 claims for G1TR-B (W49017) and G1TR-C (W49018) which are splicing  
 variants of G1TR. The G1TRs are claimed to be useful for stimulating  
 lymphocyte activity and cell death rescue. G1TR antagonists are  
 claimed to be useful for suppressing the lymphocyte activity and for  
 inducing apoptotic deletion. G1TR cDNAs and the proteins they encode  
 are also claimed to be useful for suppressing growth of tumour cells  
 over-expressing G1TR or for the treatment of refractory hodgkin's  
 disease.  
 Sequence 228 AA;  
 Query Match 4.6%; Score 11; DB 33; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 2.74e-04;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 131 fpqkthnavc 141  
 QY 143 FPGNKTNAV 153  
 RESULT 7  
 ID W37838 standard; Protein; 228 AA.  
 AC W37838.  
 DT 28-JUL-1998 (first entry)  
 DE Amino acid sequence of the mouse 312C2 T cell protein.  
 KW Mouse 312C2 T cell protein; thymus cell; spleen cell; T cell;  
 KW antigen-specific T cell proliferation; cytokine production by T-cell;  
 KW apoptosis; cancer; haematopoietic cells; lymphoid cell;  
 KW autoimmune disorders.  
 OS Mus sp.  
 PN WO9806842-A1.  
 PD 19-FEB-1998.  
 PF 14-AUG-1997; U13931.  
 PR 07-OCT-1996; US-027901.  
 PR 16-AUG-1996; US-689943.

PA (SCHE ) SCHERING CORP.  
 PI Gorman DM, Randall TD, Zlotnik A;  
 DR WPI; 98-159534/14.  
 DR N-PSDB; V19152.  
 PT Isolated 312C2 T cell gene - used to develop products for treating,  
 e.g. cancers, auto-immune disorders, transplantation rejection and  
 other T cell disorders  
 PS Claim 2; Pages 57-58; 71pp; English.  
 CC This is the amino acid sequence of the mouse 312C2 T cell protein.  
 CC The 312C2 proteins are expressed in thymus cells and are induced on  
 CC T cells and spleen cells following activation. Engagement of 312C2  
 CC stimulates proliferation of T cell clones, antigen-specific  
 CC proliferation and cytokine production by T-cells, and potentiates T  
 CC cell expansion or apoptosis. The products can be used in the  
 CC treatment of conditions associated with abnormal physiology or  
 CC development, including abnormal proliferation, e.g. cancerous  
 CC conditions or degenerative conditions. They can be used in the  
 CC regulation or development of haematopoietic cells, e.g. lymphoid cells  
 CC which affect immunological responses, e.g. autoimmune disorders.  
 SQ Sequence 228 AA;  
 Query Match 4.6%; Score 11; DB 31; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 2.74e-04;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 131 fpqkthnavc 141  
 QY 143 FPGNKTNAV 153  
 RESULT 8  
 ID W49017 standard; Protein; 294 AA.  
 AC W49017;  
 DT 29-SEP-1998 (first entry)  
 DE Mouse glucocorticoid induced TNFR-family related protein variant B.  
 KW Mouse glucocorticoid induced TNFR-family related protein; lymphocyte;  
 KW G1TR; tumour necrosis factor; apoptosis; hodgkin's disease; G1TR-B;  
 KW G1TR-C.  
 OS Mus sp.  
 PN WO9824895-A1.  
 PD 11-JUN-1998.  
 PF 08-NOV-1997; E06252.  
 PR 02-DEC-1996; GB-025074.  
 PA (PHAA ) PHARMACIA & UPJOHN SPA.  
 PI Ricciardi C;  
 DR WPI; 98-333315/29.  
 DR N-PSDB; V32774.  
 PT New isolated glucocorticoid induced TNFR related polypeptide - used  
 PT to stimulate lymphocyte activity and cell death rescue, useful to,  
 PT e.g. develop products to suppress lymphocyte activity and induce  
 PT apoptosis  
 PS Claim 14; Pages 40-41; 53pp; English.  
 CC The present claimed sequence represents a mouse glucocorticoid  
 CC induced TNFR-family related protein variant B (G1TR-B). The invention  
 CC also claims for G1TR (W49016) and G1TR-C (W49018). The G1TRs are  
 CC claimed to be useful for stimulating lymphocyte activity and cell  
 CC death rescue. G1TR antagonists are claimed to be useful for suppressing  
 CC the lymphocyte activity and for inducing apoptotic deletion. G1TR  
 CC cDNAs and the proteins they encode are also claimed to be useful for  
 CC suppressing growth of tumour cells over-expressing G1TR or for the  
 CC treatment of refractory hodgkin's disease.  
 SQ Sequence 294 AA;  
 Query Match 4.6%; Score 11; DB 33; Length 294;  
 Best Local Similarity 100.0%; Pred. No. 2.74e-04;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 131 fpqkthnavc 141  
 QY 143 FPGNKTNAV 153

RESULT 9

ID W70320 standard; Protein; 79 AA.  
AC W70320;  
DT 21-DEC-1998 (first entry)  
DE Secreted protein B0115.2  
KW Secreted protein; BD380\_1; human.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 55..67 /note= "predicted leader/signal, or transmembrane domain"  
FT Domain 29 /note= "predicted transmembrane domain is centered around residue 29"  
FT Domain 67 /note= "predicted transmembrane domain is centered around residue 67"  
FT WO9838209-A2.  
PD 03-SEP-1998.  
PT 25-FEB-1998; U03697.  
PT 24-FEB-1998; US-028724.  
PT 26-FEB-1997; US-805819.  
PA (GEM) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racie LA, Spaulding V, Treacy M;  
DR WPI: 98-481139/41.  
DR N-PSDB: V33191.  
PT New isolated polynucleotide(s) and encoded polypeptide(s) -  
PT obtained from human foetal kidney, adult colon, adult brain, foetal  
PT brain and placenta cDNA libraries.  
PS Claim 15; Page 65; 103pp; English.  
CC This is the amino acid sequence of novel human secreted protein  
CC B0115.2, as deduced from a full-length cDNA clone (see V33191)  
CC obtained from an adenocarcinoma Caco2 cDNA library. Database  
CC searching revealed some similarity between B0115.2 and some known  
CC sequences. The invention provides new isolated polynucleotides  
CC (see V33190-99), from human foetal kidney, adult colon, adult brain,  
CC foetal brain and placenta cDNA libraries, that code for secreted  
CC proteins (see W70319-27). The clones can be used for recombinant  
CC production of the polypeptides, which may have activities such as  
CC e.g. nutritional activity, cytokine and cell proliferation or  
CC differentiation activity, immunostimulant or immunosuppressive,  
CC haematopoiesis regulating activity, tissue growth activity, activin  
CC or inhibin activity, chemotactic or chemokinetic activity, activin  
CC haemostatic and thrombolytic activity, receptor/ligand activity,  
CC antiinflammatory activity, cadherin/tumour invasion suppressor  
CC activity, tumour inhibition activity, or other activities.  
CC Sequence 79 AA;

Query Match 2.9%; Score 7; DB 36; Length 79;  
Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 itvlla 27  
|||||  
QY 166 LTVVLLA 172

RESULT 10  
ID R26820 standard; Protein; 138 AA.  
AC R26820;  
DT 10-FEB-1993 (first entry)  
DE CA455 protein.  
KW Corn; stamen-specific; tassels spikelets; CA444; CA455; probe; PCR;  
KW amplify.  
OS Zea mays.  
PN WO9213957-A.  
PD 20-AUG-1992.  
PT 05-FEB-1992; E00275.  
PT 07-FEB-1991; EP-400300.  
PT 28-JUN-1991; EP-401787.  
PA (PLBZ) PLANT GENETIC SYSTEMS NV.  
PI De Beuckeleer M, Gossele V, Herdies L, Mariani C;  
DR WPI: 92-300043/36.

DR N-PSDB; Q27949.  
PT Anther-specific promoters - for control of expression of  
PT male-sterile or male fertility-restorer DNA in monocots e.g.  
PT wheat or corn  
PS Disclosure; Page 31-33; 44pp; English.  
CC The sequence given is encoded by a male flower-specific cDNA sequence  
CC isolated from corn. The cDNA sequence was isolated by using probes  
CC based on the gene core region. The cDNA sequence can be used in a  
CC foreign, chimeric DNA sequence containing a male-sterility DNA or a  
CC male-fertility restorer DNA under the transcriptional control of the  
CC promoter sequence. This vector can be used to transform the nuclear  
CC genome of a cell of a plant.  
SQ Sequence 138 AA;

Query Match 2.9%; Score 7; DB 5; Length 138;  
Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 vllavaa 17  
|||||  
QY 169 VLLAVAA 175

RESULT 11  
ID W63713 standard; Protein; 428 AA.  
AC W63713;  
DT 01-OCT-1998 (first entry)  
DE Human hIK1 protein.  
KW Intermediate conductance calcium-activated potassium channel protein 1;  
KW hIK1; human; potassium ion flux.  
OS Homo sapiens.  
PN WO9811139-A1.  
PD 19-MAR-1998.  
PT 10-SEP-1997; U16033.  
PT 17-APR-1997; US-045233.  
PT 11-SEP-1996; US-026451.  
PT 07-MAR-1997; US-040052.  
PA (ICAG-) ICAGEN INC.  
PA (UYOR-) UNIV OREGON HEALTH SCI.  
PI Adelman JP, Bond CT, Maylie J, Silvia CP;  
DR WPI: 98-207332/18.  
DR N-PSDB: V35463.  
PT DNA encoding calcium-activated potassium channel - useful in assays  
PT to identify compounds which increase or decrease potassium ion flux  
PT Claim 2; Page 117-118; 151pp; English.

CC This sequence represents the human intermediate conductance  
CC calcium-activated potassium channel protein 1 (hIK1) of the invention.  
CC The proteins of the invention are monomers of a calcium-activated  
CC potassium channel, where the monomer: (i) has a calculated molecular  
CC weight of between 40 and 80 kDa; and (ii) has a unit conductance of  
CC between 2 and 60 pS when the monomer is in the functional polymeric form  
CC of a potassium chain and is expressed in a Xenopus oocyte. Antibodies  
CC specific for the protein, and probes specific for the DNA can be used to  
CC detect the presence of the protein or DNA sequences in a sample. Host  
CC cells expression of the protein can be used in assays to identify  
CC compounds which increase or decrease the potassium ion flux through the  
CC protein. The transfected host cell can also be used for the recombinant  
CC production of the protein. The DNA sequences can also be used for  
CC determine mutations in the SK and IK genes in a computer system. The  
CC proteins encoded by the SK and IK genes can be used in a computer system  
CC for determining their three dimensional structure, which is useful for  
CC determining ligands that bind to the proteins.  
SQ Sequence 428 AA;

Query Match 2.9%; Score 7; DB 33; Length 428;  
Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 205 pgrlllg 211  
|||||  
QY 36 PGRLLLG 42

RESULT 12  
 ID W53461 standard; Protein: 498 AA.  
 AC W53461;  
 DT 27-JUL-1998 (first entry)  
 DE Human latheo protein internal reading protein sequence.  
 KW Latheo protein; fly; human; detection; blood; cerebrospinal fluid;  
 KW dopamine; Parkinson's disease; schizophrenia; depression.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 106 /note= "encoded by TTT"  
 FT Misc\_difference 111 /note= "encoded by TTT"  
 FT Misc\_difference 123 /note= "encoded by TTC"  
 FT Misc\_difference 345 /note= "encoded by TTC"  
 FT Misc\_difference 368 /note= "encoded by TTC"  
 FT Misc\_difference 409 /note= "encoded by ATT"  
 WO9810067-A1.  
 PD 12-MAR-1998.  
 PF 28-AUG-1997; U15134.  
 PR 16-JUN-1997; US-876890.  
 PR 03-SEP-1996; US-707158.  
 PA (COLD-) COLD SPRING HARBOR LAB.  
 PI Jones C, Mihalek RM, Pinto S, Tully T;  
 WPI: 98-193617/17.  
 DR N-PSDB: V23271.  
 PT DNA encoding human and fly latheo protein homologues - useful for  
 PT modulating dopamine levels in mammals  
 PS Disclosure: Fig 9A-B; 87pp; English.  
 CC This is the internal reading sequence of the human latheo protein. The  
 CC fly and human latheo proteins can modulate dopamine levels in a mammal.  
 CC Probes derived from a recombinant nucleic acid which encodes a latheo  
 CC protein are useful for the detection of latheo nucleotide sequences in  
 CC a blood or cerebral spinal fluid sample. Antibodies which bind to the  
 CC latheo protein are also useful for detection of the protein in a sample.  
 CC Agents, e.g. inhibitors, which interact with latheo protein are used to  
 CC modulate dopamine levels by altering interaction between latheo and  
 CC tyrosine hydroxylase. The ability of latheo to modulate dopamine levels  
 CC provides methods for treating a condition or disease associated with  
 CC dopamine malfunction, e.g. Parkinson's Disease, schizophrenia, and  
 CC depression.  
 CC Sequence 498 AA;  
 SQ  
 Query Match 2.9%; Score 7; DB 31; Length 498;  
 Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 184 etqllle 190  
 QY 201 ETQLLLE 207  
 |||||||  
 |||||||

RESULT 13  
 ID W53459 standard; Protein: 556 AA.  
 AC W53459;  
 DT 27-JUL-1998 (first entry)  
 DE Human latheo protein sequence.  
 KW Latheo protein; fly; human; detection; blood; cerebrospinal fluid;  
 KW dopamine; Parkinson's disease; schizophrenia; depression.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 106 /note= "encoded by TTT"  
 FT Misc\_difference 111 /note= "encoded by TTT"  
 FT Misc\_difference 123 /note= "encoded by TTC"  
 FT Misc\_difference 345 /note= "encoded by TTC"  
 FT Misc\_difference 368 /note= "encoded by TTC"  
 FT Misc\_difference 409 /note= "encoded by ATT"  
 WO9810067-A1.  
 PD 12-MAR-1998.  
 PF 28-AUG-1997; U15134.  
 PR 16-JUN-1997; US-876890.  
 PR 03-SEP-1996; US-707158.  
 PA (COLD-) COLD SPRING HARBOR LAB.  
 PI Jones C, Mihalek RM, Pinto S, Tully T;  
 WPI: 98-193617/17.  
 DR N-PSDB: V23269.  
 PT DNA encoding human and fly latheo protein homologues - useful for  
 PT modulating dopamine levels in mammals  
 PS Disclosure: Fig 9A-B; 87pp; English.  
 CC This is the internal reading sequence of the human latheo protein. The  
 CC fly and human latheo proteins can modulate dopamine levels in a mammal.  
 CC Probes derived from a recombinant nucleic acid which encodes a latheo  
 CC protein are useful for the detection of latheo nucleotide sequences in  
 CC a blood or cerebral spinal fluid sample. Antibodies which bind to the  
 CC latheo protein are also useful for detection of the protein in a sample.  
 CC Agents, e.g. inhibitors, which interact with latheo protein are used to  
 CC modulate dopamine levels by altering interaction between latheo and  
 CC tyrosine hydroxylase. The ability of latheo to modulate dopamine levels  
 CC provides methods for treating a condition or disease associated with  
 CC dopamine malfunction, e.g. Parkinson's Disease, schizophrenia, and  
 CC depression.  
 CC Sequence 556 AA;  
 SQ  
 Query Match 2.9%; Score 7; DB 31; Length 556;  
 Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 242 etqllle 248  
 QY 201 ETQLLLE 207  
 |||||||  
 |||||||

PT DNA encoding human and fly latheo protein homologues - useful for  
 PT modulating dopamine levels in mammals  
 PS Claim 15; Fig 6; 87pp; English.  
 CC This represents a human latheo protein. The fly and human latheo  
 CC proteins can modulate dopamine levels in a mammal. Probes derived from a  
 CC recombinant nucleic acid which encodes a latheo protein are useful for  
 CC the detection of latheo nucleotide sequences in a blood or cerebral  
 CC spinal fluid sample. Antibodies which bind to the latheo protein are  
 CC also useful for detection of the protein in a sample. Agents, e.g.  
 CC inhibitors, which interact with latheo protein are used to modulate  
 CC dopamine levels by altering interaction between latheo and tyrosine  
 CC hydroxylase. The ability of latheo to modulate dopamine levels provides  
 CC methods for treating a condition or disease associated with dopamine  
 CC malfunction, e.g. Parkinson's Disease, schizophrenia, and depression.  
 CC Sequence 556 AA;  
 SQ  
 Query Match 2.9%; Score 7; DB 31; Length 556;  
 Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 242 etqllle 248  
 QY 201 ETQLLLE 207  
 |||||||  
 |||||||

RESULT 14  
 ID W25638 standard; Protein: 713 AA.  
 AC W25638;  
 DT 03-NOV-1997 (first entry)  
 DE Human cadherin-13.  
 KW Human; cadherin; rat; calcium-dependent cell adhesion protein;  
 KW superfamily; cytoskeleton; eatenin; cancer.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 703 /note= "Encoded by CTC"  
 FT US5646250-A.  
 PN 08-JUL-1997.  
 PD 17-APR-1992; 872643.  
 PR 19-APR-1993; US-049460.  
 PR 17-APR-1992; US-872643.  
 PR 01-NOV-1994; US-332638.  
 PA (DOHE-) DOHENY EYE INST.  
 PI Suzuki S;  
 WPI: 97-362997/33.  
 DR N-PSDB: T85405.  
 PT Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion  
 PS Claim 1: Column 99-102; 56pp; English.  
 CC This sequence represents human cadherin-13. The invention specifically  
 CC provides details of human cadherin-5, -8, -11, -12 and -13, and rat  
 CC cadherin-8, -5, -11 and -13. Cadherins are calcium-dependent cell  
 CC adhesion proteins. They are glycosylated integral membrane proteins  
 CC that have an N-terminal extracellular domain that determines binding  
 CC specificity, a hydrophobic membrane spanning region and a C-terminal  
 CC cytoplasmic domain, which is highly conserved among members of the  
 CC superfamily. The C-terminal domain interacts with the cytoskeleton  
 CC through eatenins and other cytoskeleton-associated proteins. The  
 CC novel cadherin proteins may be used in the analysis of the role of  
 CC cadherins in various cancers. Sequence analysis of the cadherin  
 CC proteins also allows investigation of the structure and function of  
 CC cadherin. The cadherin proteins may be isolated by using anti-cadherin  
 CC antibodies. These antibodies may also be used to modulate the activity  
 CC of cadherin and to determine the tissue specific distribution of cadherin  
 CC proteins. Each subclass of cadherins has a unique tissue distribution  
 CC pattern.  
 CC Sequence 713 AA;  
 SQ  
 Query Match 2.9%; Score 7; DB 23; Length 713;  
 Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 16 vllllsa 22  
 QY 201 ETQLLLE 207  
 |||||||  
 |||||||

QY 177 VLLTSA 183

## RESULT 15

ID W13136 standard; Protein; 713 AA.  
 AC W13136;  
 DT 14-MAY-1997 (first entry)  
 DE Full length human cadherin-13.  
 KW Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;  
 KW brain; human; antibody; purification; determination;  
 KW tissue expression; binding antagonist; calcium ion.  
 OS Homo sapiens.  
 PN US5597725-A.  
 PD 28-JAN-1997.  
 PF 17-APR-1992; 872643.  
 PR 17-APR-1992; US-872643.  
 PR 19-APR-1993; US-049460.  
 PR 26-JAN-1994; US-188228.  
 PA (DOHE-) DOHENY EYE INST.  
 AU Suzuki S;  
 MI WPI: 97-108328/10.  
 DR N-PSDB: T61927.  
 PT Antibodies to cadherin proteins - useful as cadherin antagonists,  
 etc.  
 PS Example 2; Columns 107-112; 59pp; English.  
 CC The present sequence is full length human cadherin-13, which  
 is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA  
 was isolated from a foetal brain cDNA library, using probes based  
 on homologous rat cadherin cDNA.  
 CC Antibodies or fragments that specifically bind the human cadherin  
 can be used to purify the cadherin, determine its tissue expression  
 and antagonise its ligand/antiligand binding activities.  
 SQ Sequence 713 AA;

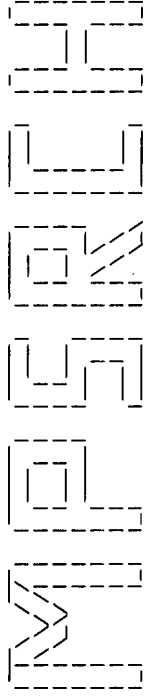
Query Match 2.9%; Score 7; DB 21; Length 713;  
 Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 vllltsa 22

QY 177 VLLTSA 183

Search completed: Mon Jul 12 12:12:44 1999  
 Job time : 27 secs.

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
on: Mon Jul 12 12:10:58 1999; MasPar time 4.30 Seconds  
Modular output not generated. 538.908 Million cell updates/sec

Title: >US-08-911-423-2  
Description: (1-228) from US08911423.pep  
Perfect Score: 228  
Sequence: 1 MGANWMLGVSMCLVLDLQ.....PEERGETEKCHLGRWP 228

Scoring table: TABLE uniprotatable  
Gap 60  
Searched: 106580 seqs, 10152877 residues  
Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1  
Statistics: Mean 2.565; Variance 0.611; scale 4.197

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	6	2.6	84	3	PCT-US94-0 Sequence 4, Applicatio	4.45e+01
2	6	2.6	84	3	PCT-US91-0 Sequence 4, Applicatio	4.45e+01
3	6	2.6	84	2	US-08-485- Sequence 4, Applicatio	4.45e+01
4	6	2.6	84	2	US-08-407- Sequence 4, Applicatio	4.45e+01
5	6	2.6	84	1	US-07-853- Sequence 4, Applicatio	4.45e+01
6	6	2.6	84	1	US-08-184- Sequence 4, Applicatio	4.45e+01
7	6	2.6	84	1	US-07-681- Sequence 4, Applicatio	4.45e+01
8	6	2.6	223	1	US-08-483- Sequence 13, Applicati	4.45e+01
9	6	2.6	223	1	US-08-472- Sequence 13, Applicati	4.45e+01
10	6	2.6	223	1	US-08-278- Sequence 13, Applicati	4.45e+01
11	6	2.6	223	2	US-08-487- Sequence 13, Applicatio	4.45e+01
12	6	2.6	295	2	US-08-454- Sequence 7, Applicatio	4.45e+01
13	6	2.6	326	4	5268278-3 Patent No. 5268278.	4.45e+01
14	6	2.6	342	2	US-08-742- Sequence 2, Applicatio	4.45e+01
15	6	2.6	481	3	PCT-US93-0 Sequence 9, Applicatio	4.45e+01
16	6	2.6	496	3	PCT-US94-0 Sequence 15, Applicatio	4.45e+01
17	6	2.6	504	3	PCT-US91-0 Sequence 6, Applicatio	4.45e+01
18	6	2.6	504	3	PCT-US91-0 Sequence 6, Applicatio	4.45e+01
19	6	2.6	504	2	US-08-485- Sequence 6, Applicatio	4.45e+01
20	6	2.6	504	2	US-08-407- Sequence 6, Applicatio	4.45e+01
21	6	2.6	504	1	US-07-681- Sequence 6, Applicatio	4.45e+01
22	6	2.6	504	1	US-08-184- Sequence 6, Applicatio	4.45e+01
23	6	2.6	504	1	US-07-853- Sequence 6, Applicatio	4.45e+01

97 5 2.2 562 4 5185259-3 Patent No. 5185259. 5.57e+02  
98 5 2.2 652 4 5258288-4 Patent No. 5258288. 5.57e+02  
99 5 2.2 1280 4 5206352-4 Patent No. 5206352. 5.57e+02  
100 5 2.2 1365 4 5194600-4 Patent No. 5194600. 5.57e+02

## ALIGNMENTS

RESULT 1  
ID PCT-US94-04174-4 STANDARD; PRT; 84 AA.

XX xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 4, Application PC/TUS9404174

Sequence 4, Application PC/TUS9404174

GENERAL INFORMATION:

APPLICANT: Reyes, Gregory

APPLICANT: Kim, Jungsuh P.

APPLICANT: Moeckli, Randolph

TITLE OF INVENTION: HCV Immunodiagnostic Antigens and Antibodies

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Peter J. Dehlinger

STREET: 350 Cambridge Avenue, Suite 300

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/04174

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION NUMBER: US 505,611

FILING DATE: 06-APR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Fabian, Gary R.

REGISTRATION NUMBER: 33,875

REFERENCE/DOCKET NUMBER: 4600-0107

TELEPHONE: 415-324-0980

TELEFAX: 415-324-0980

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 84 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 84 AA; 9408 MW; 41094 CN;

Query Match 2.6%; Score 6; DB 3; Length 84;

Best Local Similarity 100.0%; Pred. No. 4.45e+01;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 78 PGORVE 83

| | | | |

QY 84 PGORVE 89

RESULT 2

ID PCT-US91-02370-4 STANDARD; PRT; 84 AA.

XX xxxxxx

XX

XX

XX

XX

Sequence 4, Application US/08485500

Sequence 4, Application US/08485500

Patent No. 5843639

GENERAL INFORMATION:

APPLICANT: Reyes, Gregory

APPLICANT: Kim, Jungsuh P.

APPLICANT: Moeckli, Randolph

APPLICANT: Simonsen, Christian C.

TITLE OF INVENTION: Hepatitis C Virus Epitopes

XX Sequence 4, Application PC/TUS9102370

XX Sequence 4, Application PC/TUS9102370

XX GENERAL INFORMATION:

XX APPLICANT: Reyes, Gregory

XX APPLICANT: Kim, Jungsuh P.

XX APPLICANT: Moeckli, Randolph

XX APPLICANT: Simonsen, Christian C.

XX TITLE OF INVENTION: Hepatitis C Virus Epitopes

XX NUMBER OF SEQUENCES: 26

XX CORRESPONDENCE ADDRESS:

XX ADDRESSEE: Peter J. Dehlinger

XX STREET: P.O. BOX 60850

XX CITY: Palo Alto

XX STATE: CA

XX COUNTRY: USA

XX ZIP: 94306

XX COMPUTER READABLE FORM:

XX MEDIUM TYPE: Floppy disk

XX COMPUTER: IBM PC compatible

XX OPERATING SYSTEM: PC-DOS/MS-DOS

XX SOFTWARE: PatentIn Release #1.0, Version #1.25

XX CURRENT APPLICATION DATA:

XX APPLICATION NUMBER: PCT/US91/02370

XX FILING DATE: 19910405

XX CLASSIFICATION: 435.5

XX PRIOR APPLICATION NUMBER: US 505,611

XX FILING DATE: 06-APR-1990

XX PRIOR APPLICATION DATA:

XX APPLICATION NUMBER: US 594,854

XX FILING DATE: 09-OCT-1990

XX ATTORNEY/AGENT INFORMATION:

XX NAME: Fabian, Gary R.

XX REGISTRATION NUMBER: 33,875

XX REFERENCE/DOCKET NUMBER: 4600-076.41

XX TELECOMMUNICATION INFORMATION:

XX TELEPHONE: (415) 323-8302

XX INFORMATION FOR SEQ ID NO: 4:

XX SEQUENCE CHARACTERISTICS:

XX LENGTH: 84 amino acids

XX TYPE: AMINO ACID

XX TOPOLOGY: linear

XX MOLECULE TYPE: protein

XX SEQUENCE 84 AA; 9408 MW; 41094 CN;

Query Match 2.6%; Score 6; DB 3; Length 84;

Best Local Similarity 100.0%; Pred. No. 4.45e+01;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 78 PGORVE 83

| | | | |

QY 84 PGORVE 89

RESULT 3

ID US-08-485-500-4 STANDARD; PRT; 84 AA.

XX xxxxxx

XX

XX

XX

XX

XX

Sequence 4, Application US/08485500

Sequence 4, Application US/08485500

Patent No. 5843639

GENERAL INFORMATION:

APPLICANT: Reyes, Gregory

APPLICANT: Kim, Jungsuh P.

APPLICANT: Moeckli, Randolph

APPLICANT: Simonsen, Christian C.

TITLE OF INVENTION: Hepatitis C Virus Epitopes



CC FILING DATE: 06-APR-1990  
CC APPLICATION NUMBER: US 07/594,854  
CC FILING DATE: 09-OCT-1990  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Dehlinger, Peter J.  
CC REGISTRATION NUMBER: 28,006  
CC REFERENCE/DOCKET NUMBER: 4600-0076.22  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 415-324-0880  
CC TELEFAX: 415-324-0960  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 84 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 84 AA; 9408 MW; 41094 CN;

Query Match 2.6%; Score 6; DB 1; Length 84;  
Best Local Similarity 100.0%; Pred. No. 4.45e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 78 PGORVE 83  
|||||  
QY 84 PGORVE 89

RESULT 6  
ID US-08-184-236-4 STANDARD; PRT; 84 AA.

XXXXXX

Sequence 4, Application US/08184236

Sequence 4, Application US/08184236  
Patent No. 5538865  
GENERAL INFORMATION:  
CC APPLICANT: Reyes, Gregory  
CC APPLICANT: Kim, Jungsuh P.  
CC APPLICANT: Moeckli, Randolph  
CC TITLE OF INVENTION: Hepatitis C Virus Epitopes  
CC NUMBER OF SEQUENCES: 10  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Dehlinger & Associates  
CC STREET: P.O. BOX 60850  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: US  
CC ZIP: 94306  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.24  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/184,236  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/505,611  
CC FILING DATE: 06-APR-1990  
CC APPLICATION NUMBER: US 07/594,854  
CC FILING DATE: 09-OCT-1990  
CC APPLICATION NUMBER: US 07/853,985  
CC FILING DATE: 20-MAR-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Fabian, Gary R.  
CC REGISTRATION NUMBER: 33,875  
CC REFERENCE/DOCKET NUMBER: 4600-0113  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 415-324-0880

CC TELEFAX: 415-324-0960  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 84 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 84 AA; 9408 MW; 41094 CN;

Query Match 2.6%; Score 6; DB 1; Length 84;  
Best Local Similarity 100.0%; Pred. No. 4.45e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 78 PGORVE 83  
|||||  
QY 84 PGORVE 89

RESULT 7  
ID US-07-681-703B-4 STANDARD; PRT; 84 AA.

XXXXXX

Sequence 4, Application US/07681703B

Sequence 4, Application US/07681703B  
Patent No. 5443965  
GENERAL INFORMATION:  
CC APPLICANT: Reyes, Gregory  
CC APPLICANT: Kim, Jungsuh P.  
CC APPLICANT: Moeckli, Randolph  
CC TITLE OF INVENTION: Hepatitis C Virus Epitopes  
CC NUMBER OF SEQUENCES: 55  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Dehlinger & Associates  
CC STREET: 350 Cambridge Ave., Suite 250  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94306  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/681,703B  
CC FILING DATE: 05-APR-1991  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 505,611  
CC FILING DATE: 06-APR-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 594,854  
CC FILING DATE: 09-OCT-1990  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Fabian, Gary R.  
CC REGISTRATION NUMBER: 33,875  
CC REFERENCE/DOCKET NUMBER: 4600-076.21  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 324-0880  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 84 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 84 AA; 9408 MW; 41094 CN;

Query Match 2.6%; Score 6; DB 1; Length 84;  
Best Local Similarity 100.0%; Pred. No. 4.45e+01;



DE Sequence 13, Application US/08278091  
XX Sequence 13, Application US/08278091  
CC Patent No. 5506139  
CC GENERAL INFORMATION:  
CC APPLICANT: LOOSMORE, Sheena M  
CC APPLICANT: YANG, Yan-Ping  
CC APPLICANT: CHONG, Pele  
CC APPLICANT: OOMEN, Raymond P.  
CC APPLICANT: KLEIN, Michel H.  
CC TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
CC TITLE OF INVENTION: Reduced Protease Activity  
CC NUMBER OF SEQUENCES: 23  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Sim & McBurney  
CC STREET: Suite 701, 330 University Avenue  
CC CITY: Toronto  
CC STATE: Ontario  
CC COUNTRY: Canada  
CC ZIP: M5G 1R7  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/278,091  
CC FILING DATE: 21-JUL-1994  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Stewart, Michael I.  
CC REGISTRATION NUMBER: 24,973  
CC REFERENCE/DOCKET NUMBER: 1038-371  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (416) 595-1155  
CC TELEFAX: (416) 595-1163  
CC INFORMATION FOR SEQ ID NO: 13:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 223 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC SEQUENCE 223 AA; 23060 MW; 248965 CN;  
SQ  
Query Match 2.6%; Score 6; DB 1; Length 223;  
Best Local Similarity 100.0%; Pred. No. 4.45e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 41 GSGNNT 46  
Qy 37 GSGNNT 42  
RESULT 11  
ID US-08-487-167-13 STANDARD; PRT; 223 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 13, Application US/08487167  
XX  
CC Sequence 13, Application US/08487167  
CC Patent No. 5869302  
CC GENERAL INFORMATION:  
CC APPLICANT: LOOSMORE, Sheena M.  
CC APPLICANT: YANG, Yan-Ping  
CC APPLICANT: CHONG, Pele  
CC APPLICANT: OOMEN, Raymond P.  
CC APPLICANT: KLEIN, Michel H.  
CC TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
CC TITLE OF INVENTION: Reduced Protease Activity  
CC NUMBER OF SEQUENCES: 23

CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Sim & McBurney  
CC STREET: Suite 701, 330 University Avenue  
CC CITY: Toronto  
CC STATE: Ontario  
CC COUNTRY: Canada  
CC ZIP: M5G 1R7  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/487,167  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC PRIOR APPLICATION NUMBER: US 08/296,149  
CC FILING DATE: 26-AUG-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/278,091  
CC FILING DATE: 21-JUL-1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Stewart, Michael I.  
CC REGISTRATION NUMBER: 24,973  
CC REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (416) 595-1155  
CC TELEFAX: (416) 595-1163  
CC INFORMATION FOR SEQ ID NO: 13:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 223 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC SEQUENCE 223 AA; 23060 MW; 248965 CN;  
SQ  
Query Match 2.6%; Score 6; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 4.45e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 41 GSGNNT 46  
Qy 37 GSGNNT 42  
RESULT 12  
ID US-08-454-267-7 STANDARD; PRT; 295 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 7, Application US/08454267  
XX  
CC Sequence 7, Application US/08454267  
CC Patent No. 5843739  
CC GENERAL INFORMATION:  
CC APPLICANT: SLABAS, ANTONI R.  
CC APPLICANT: BROWN, ADRIAN P.  
CC TITLE OF INVENTION: DNA ENCODING 2-ACYLTRANSFERASES  
CC NUMBER OF SEQUENCES: 7  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
CC STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
CC CITY: WASHINGTON  
CC STATE: DC  
CC COUNTRY: US  
CC ZIP: 20005-3934  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/454,267  
CC FILING DATE: 08-JUN-1995  
CC CLASSIFICATION: 800  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/GB93/02528  
CC FILING DATE: 10-DEC-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: REED, GRANT E.  
CC REGISTRATION NUMBER: P-41,264  
CC REFERENCE/DOCKET NUMBER: 0623.0310000/JAG/GER  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202) 371-2600  
CC TELEFAX: (202) 371-2540  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 295 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 295 AA; 32665 MW; 456556 CN;

Query Match 2.6%; Score 6; DB 2; Length 295;  
Best Local Similarity 100.0%; Pred. No. 4.45e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 163 GQPSV 168  
|||||  
QY 19 GQPSV 24

RESULT 13  
ID 5268278-3 STANDARD; PRT: 353 AA.

XX  
AC xxxxxx  
DT 01-JAN-1900

XX Patent No. 5268278.

XX Patent No. 5268278  
CC APPLICANT: CANOSI, UMBERTO; DE FAZIO, GABRIELE; VILLA, STEFANO;  
CC DONINI, SILVIA  
CC TITLE OF INVENTION: GENETIC EXPRESSION OF SOMATOSTATIN AS  
CC HYBRID POLYPEPTIDE  
CC NUMBER OF SEQUENCES: 4

CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/193,202  
CC FILING DATE: 09-MAY-1988  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 717,444  
CC FILING DATE: 29-MAR-1985  
CC SEQ ID NO: 3:  
CC LENGTH: 326  
CC SEQUENCE 353 AA; 39277 MW; 696100 CN;

Query Match 2.6%; Score 6; DB 4; Length 326;  
Best Local Similarity 100.0%; Pred. No. 4.45e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 163 QLSAED 168  
|||||  
QY 197 QLSAED 202

RESULT 14  
ID US-08-742-011-2 STANDARD; PRT: 342 AA.

XX  
AC xxxxxx  
DT  
XX

DE Sequence 2, Application US/08742011  
XX Sequence 2, Application US/08742011  
CC Patent No. 5824504  
CC GENERAL INFORMATION:  
CC APPLICANT: Elshourbagy, Nabil A.  
CC APPLICANT: Bergsma, Derk J.  
CC APPLICANT: Ellis, Catherine E.  
CC TITLE OF INVENTION: Human 7-Transmembrane Receptor  
CC NUMBER OF SEQUENCES: 2  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Smithkline Beecham Corporation  
CC STREET: 709 Swedeland Road  
CC CITY: King of Prussia  
CC STATE: Pennsylvania  
CC COUNTRY: USA  
CC ZIP: 19406-2799  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/742,011  
CC FILING DATE:  
CC CLASSIFICATION: 530  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Han, William T.  
CC REGISTRATION NUMBER: 34,344  
CC REFERENCE/DOCKET NUMBER: ATG50020P  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 610-270-5219  
CC TELEFAX: 610-270-5090  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 342 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 342 AA; 39280 MW; 683920 CN;

Query Match 2.6%; Score 6; DB 2; Length 342;  
Best Local Similarity 100.0%; Pred. No. 4.45e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 233 IFLVMA 238  
|||||  
QY 157 IFLVMA 162

RESULT 15  
ID PCT-US93-07213-9 STANDARD; PRT: 481 AA.

XX  
AC xxxxxx

XX

XX Sequence 9, Application PC/TUS9307213

XX Sequence 9, Application PC/TUS9307213  
CC GENERAL INFORMATION:

CC APPLICANT: The Government of the United States of  
CC APPLICANT: America, as represented by The Secretary  
CC TITLE OF INVENTION: NUCLEIC ACIDS AND AMINO ACID  
CC TITLE OF INVENTION: SEQUENCES FOR A STRONGLY IMMUNOREACTIVE PROTEIN ENCODED  
CC TITLE OF INVENTION: HUMAN HERPESVIRUS 6 STRAIN 229  
CC NUMBER OF SEQUENCES: 16  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: The Government of the United States of  
CC ADDRESSEE: America, as represented by The Secretary  
CC STREET: 6011 Executive Blvd., Suite 325  
CC CITY: Rockville  
CC STATE: Maryland

CC COUNTRY: United States of America  
 CC ZIP: 20852  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: PCT/US93/07213  
 CC FILING DATE: 19930730  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/923,743  
 CC FILING DATE: 31-JULY-1992  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (301) 496-7056  
 CC TELEFAX: (301) 402-0220  
 CC TELEX: None  
 CC INFORMATION FOR SEQ ID NO: 9:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 481 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 481 AA; 52789 MW; 1212833 CN;  
 Query Match 2.6%; Score 6; DB 3; Length 481;  
 Best Local Similarity 100.0%; Pred. No. 4.45e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 109 GORVES 114  
 QY 85 GORVES 90

Search completed: Mon Jul 12 12:11:13 1999  
 Job time : 15 secs.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MPPerch_pp      protein - protein database search, using Smith-Waterman algorithm
on:              Mon Jul 12 12:14:20 1999;      MasPar time 17.63 seconds
                  746.201 Million cell updates/sec
Tabular output not generated.

```

```
>US-08-911-423-4
Title:
Description: (1-241) from US08911423.pep
Perfect Score: 241
Sequence: 1 MAQHGAMGAFALCGIALIC.....EEEGERSAEKGRIGDLWV 241
```

Scoring table: TABLE unitprototable  
Gap 60

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database:

```
sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_orga
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unccla
13:sp_vertebrate 14:sp_virus
```

Statistics: Mean 3.502; Variance 0.456; scale 7.678

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	11	4.6	228	Q35714	GLUCOCORTICOID INDUCED	1.44e+08
2	8	3.3	241	Q57793	HYPOTHETICAL 27.5 KD P	4.53e-02
3	3	3.3	245	Q14596	FOLATE BINDING PROTEIN	4.53e-02
4	8	3.3	438	Q39493	GLYCOPROTEIN M.	4.53e-02
5	8	3.3	504	Q75850	WUGSC-H.DJ0751H13.3 PR	4.53e-02
6	8	3.3	509	Q05457	HYPOTHETICAL 53.2 KD P	4.52e-02
7	7	2.9	120	Q33793	VALINE DEHYDROGENASE (	3.41e+00
8	7	2.9	133	Q38336	133AA LONG HYPOTHETICA	3.41e+00
9	7	2.9	172	Q34470	OUTER MEMBRANE PROTEIN	3.41e+00
10	7	2.9	206	Q32017	YQII PROTEIN.	3.41e+00
11	7	2.9	222	Q14569	MATCH TO AA456453.	3.41e+00
12	7	2.9	232	Q780943	T-CELL SURFACE GLYCOPR	3.41e+00
13	7	2.9	256	Q31618	YBUT PROTEIN.	3.41e+00
14	7	2.9	264	Q77255	EG-22E5.6 PROTEIN.	3.41e+00
15	7	2.9	332	Q37402	NADH DEHYDROGENASE, SU	3.41e+00
16	7	2.9	333	Q29422	T-CELL SURFACE GLYCOPR	3.41e+00
17	7	2.9	333	Q28565	T-CELL SURFACE GLYCOPR	3.41e+00
18	7	2.9	350	Q28052	ORNITHINE CYCLODEAMINA	3.41e+00
19	7	2.9	352	Q35557	HYPOTHETICAL 38.5 KD P	3.41e+00
20	7	2.9	371	Q91984	C50B6.10 PROTEIN.	3.41e+00

94 6 2.5 566 14 Q82774 HAEMAGGLUTININ PRECURS 1.58e+02  
 95 6 2.5 566 14 P88836 HAEMAGGLUTININ HA (FRA 1.58e+02  
 96 6 2.5 566 14 Q26937 HEAT SHOCK PROTEIN 70. 1.58e+02  
 97 6 2.5 787 10 Q41583 INITIATION FACTOR (ISO 1.58e+02  
 98 6 2.5 3015 14 Q92532 POLYPROTEIN. 1.58e+02  
 99 6 2.5 3898 14 Q92872 POLYPROTEIN. 1.58e+02  
 100 6 2.5 3898 14 Q92365 POLYPROTEIN. 1.58e+02

## ALIGNMENTS

RESULT 1  
 ID Q35714 PRELIMINARY; PRT; 228 AA.  
 AC Q35714;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOCORTICOID INDUCED TNFR FAMILY RELATED PROTEIN PRECURSOR.  
 GN GTR.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 RN SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=C3H/HEN;  
 RC MEDLINE: 97322352.  
 RX NOCENTINI G., GIUNCHI L., RONCHETTI S., KRAUSZ L.T., BARTOLI A.,  
 RA MORACA R., MIGLIORATI G., RICCARDI C.;  
 RT "A new member of the tumor necrosis factor/nerve growth factor  
 RT receptor family inhibits T cell receptor-induced apoptosis.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:6216-6221(1997).  
 DR EMBL: U82534; G2228584; -;  
 MGD; MGI:894675; GTR.  
 KW SIGNAL.  
 FT CHAIN 1 19 POTENTIAL.  
 FT CHAIN 20 228 GLUCOCORTICOID INDUCED TNFR FAMILY  
 FT RELATED PROTEIN.  
 SQ SEQUENCE 228 AA; 25334 MW; BA433757 CRC32;

Query Match 4.6%; Score 11; DB 11; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 1.44e-08;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 131 FPGKTHNAV 141  
 QY 143 FPGKTHNAV 153

RESULT 2  
 AC Q67793 PRELIMINARY; PRT; 241 AA.  
 AC Q67793;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 27.5 KD PROTEIN.  
 GN AQ\_1986.  
 OS AQUIFEX AEOLICUS.  
 OC BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=VF5;  
 RC MEDLINE: 98196666.  
 RX DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus.";  
 RL NATURE 392:353-358(1998).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=VF5;  
 RC DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,

RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF000766; G2984233; -;  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 241 AA; 27514 MW; A94A054B CRC32;

Query Match 3.3%; Score 8; DB 2; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 4.53e-02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 38 ALLCALS 45  
 QY 17 ALLCALS 24

RESULT 3  
 ID Q14596 PRELIMINARY; PRT; 245 AA.  
 AC Q14596;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE FOLATE BINDING PROTEIN.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 RN CATARRHINI; HOMINIDAE; HOMO.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SALIVARY GLAND;  
 RA VERMA R.S., ELWOOD P.C.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AF000380; G2565194; -;  
 SQ SEQUENCE 245 AA; 26268 MW; 4FF90C3F CRC32;

Query Match 3.3%; Score 8; DB 4; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 4.53e-02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 177 LTSQQLGL 184  
 QY 180 LTSQQLGL 187

RESULT 4  
 ID Q39493 PRELIMINARY; PRT; 438 AA.  
 AC Q39493;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLYCOPROTEIN M.  
 GN UL10.  
 OS BOVINE HERPESVIRUS TYPE 1.  
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;  
 OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=JURA;  
 RC MEDLINE: 95313343.  
 RX VLCEK C., BENES V., LU Z., KUTISH G.F., PACES V., ROCK D.,  
 RA LETCHWORTH G.J., SCHWYZER M.;  
 RT "Nucleotide sequence analysis of a 30-kb region of the bovine  
 RT herpesvirus 1 genome which exhibits a colinear gene arrangement with  
 RT the UL21 to UL4 genes of herpes simplex virus.";  
 RL VIROLOGY 210:100-108(1995).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=JURA;  
 RC SCHWYZER M., PACES V., LETCHWORTH G.J., MISRA V., BUHK H.J.,  
 RA LOWERY D.E., SIMARD C., BELLO L.J., THIRY E., VLCEK C.;  
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: AJ004801; E1187334; -;  
 SQ SEQUENCE 438 AA; 45517 MW; 933A012A CRC32;

Query Match 3.3%; Score 8; DB 14; Length 438;

Best Local Similarity 100.0%; Pred. No. 4.53e-02; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Length 504 AA.

Db 93 VLLAVAA 100  
QY 168 VLLAVAA 175

RESULT 5  
ID O75850 PRELIMINARY; PRT; 504 AA.  
AC O75850;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DE WUGSC:H.DJ0751H13.3 PROTEIN.  
OS HOMO SAPIENS (HUMAN)  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
[1]  
SEQUENCE FROM N.A.  
LEONARD S., GRAVES T., STROMWATT C.;  
"The sequence of Homo sapiens PAC clone DJ0751H13.";  
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA WATERSTON R.H.;  
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA WATERSTON R.;  
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AC004877; G3638956; -;  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 12.  
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.  
SQ SEQUENCE 504 AA; 55490 MW; 74FD7A10 CRC32;

Query Match 3.3%; Score 8; DB 4; Length 504;  
Best Local Similarity 100.0%; Pred. No. 4.53e-02; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Length 504 AA.

Db 317 GPGCGPG 324  
QY 30 GPGCGPG 37

RESULT 6  
ID O05457 PRELIMINARY; PRT; 509 AA.  
AC O05457;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DE HYPOTHETICAL 53.2 KD PROTEIN.  
GN MTCY15F10.25.  
OS MYCOBACTERIUM TUBERCULOSIS.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA OLIVER K., HARRIS D.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA COLE S.T., BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE: 96181548.  
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L., JACOBS W.R. JR.,  
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
RA COLE S.T.;

\*An integrated map of the genome of the tubercle bacillus,  
RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium  
leprae.\*;  
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
DR EMBL: Z94121; E312290; -;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 509 AA; 53278 MW; 04302F67 CRC32;

Query Match 3.3%; Score 8; DB 2; Length 509;  
Best Local Similarity 100.0%; Pred. No. 4.53e-02; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Length 509 AA.

Db 205 VLLAVAA 212  
QY 168 VLLAVAA 175

RESULT 7  
ID Q53783 PRELIMINARY; PRT; 120 AA.  
AC Q53783;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)  
DE VALINE DEHYDROGENASE (VDH) (FRAGMENT).  
OS STREPTOMYCES AMBOFACIENS.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
OC ACTINOMYCETALES; STREPTOMICINAE; STREPTOMICETACEAE; STREPTOMYCES.  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95014047.  
RA TANG L., ZHANG Y.X., HUTCHINSON C.R.;  
RT "Amino acid catabolism and antibiotic synthesis: valine is a source of  
RT precursors for macrolide biosynthesis in Streptomyces ambofaciens and  
RT Streptomyces fradiae.";  
RL J. BACTERIOL. 176:6107-6119(1994).  
DR EMBL: L33871; G688448; -;  
FT NON\_TER 120 120  
SQ SEQUENCE 120 AA; 13690 MW; AEA28DBB CRC32;

Query Match 2.9%; Score 7; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 3.41e+00; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Length 120 AA.

Db 24 FPEERG 30  
QY 220 FPEERG 226

RESULT 8  
ID O58336 PRELIMINARY; PRT; 133 AA.  
AC O58336;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)  
DE 133AA LONG HYPOTHETICAL PROTEIN.  
GN PH0581.  
OS PYROCOCCLUS HORIKOSHII.  
OC ARCHAEA; EUARYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCLUS.  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-OT3;  
RX MEDLINE: 98344137.  
RA KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSUYAMA A., NAGAI Y.,  
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
RA KIKUCHI H.;  
RT "Complete Sequence and Gene Organization of the Genome of a  
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA RES. 5:55-76(1998).  
DR EMBL: AF000002; D1030613; -;  
SQ SEQUENCE 133 AA; 15160 MW; 0DF44E0F CRC32;

Query Match 2.9%; Score 7; DB 1; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 3.41e+00;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 103 GRLGTT 109  
 QY 37 GRLGTT 43  
 |||||

RESULT 9  
 ID Q54470 PRELIMINARY; PRT; 172 AA.  
 AC Q54470;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE OUTER MEMBRANE PROTEIN PRECURSOR.  
 GN OMP4.  
 OS SERRATIA MARCESCENS.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
 OC SERRATIA.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N28B;  
 RX MEDLINE; 96036211.  
 RA GUASCH J.F., FERRER S., ENFEDAQUE J., VIEJO M.B., REGUE M.;  
 RT "A 17 kDa outer-membrane protein (Omp4) from Serratia marcescens  
 RT confers partial resistance to bacteriocin 28b when expressed in  
 RT Escherichia coli.";  
 RL MICROBIOLOGY 141:2535-2542(1995).  
 DR EMBL; 237157; G587548; -  
 KW SIGNAL; OUTER MEMBRANE.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 172 OUTER MEMBRANE PROTEIN.  
 SQ SEQUENCE 172 AA; 18434 MW; 7C38C7F8 CRC32;

Query Match 2.9%; Score 7; DB 2; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 3.41e+00;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 AVAACVL 15  
 QY 172 AVAACVL 178  
 |||||

RESULT 10  
 ID Q32017 PRELIMINARY; PRT; 206 AA.  
 AC Q32017;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE YQII PROTEIN.  
 GN YQII.  
 OS BACILLUS SUBTILIS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
 OC BACILLUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE; 98044033.  
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,  
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,  
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,  
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,  
 RA ENTIAN K.D., ERINGTON J., FABRET C., FERRARI E., FOULGER D.,  
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,  
 RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,  
 RA GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,  
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,  
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,  
 RA KOBAYASHI Y., KOETTER P., KONIGSTEIN G., KROGH S., KUMANO M.,

RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,  
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,  
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,  
 RA NOONE D., O'REILLY M., OGAWA K., OGAWA B., OUDEGA B., PARK S.H.,  
 RA PARRO V., POHL T.M., PORTELELE D., PORWOLLIK S., PRESCOTT A.M.,  
 RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,  
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADATE Y.,  
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,  
 RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,  
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,  
 RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,  
 RA VIARI A., WAMBUYT R., WEDLER E., WEDLER H., WEITZENEGGER T.,  
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,  
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL NATURE 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 DR EMBL; 299116; E1185687; -  
 SQ SEQUENCE 206 AA; 22235 MW; 5B419CAF CRC32;

Query Match 2.9%; Score 7; DB 2; Length 206;  
 Best Local Similarity 100.0%; Pred. No. 3.41e+00;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 ALCGLAL 15  
 QY 12 ALCGLAL 18  
 |||||

RESULT 11  
 ID O14569 PRELIMINARY; PRT; 222 AA.  
 AC O14569;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE MATCH TO AA456453.  
 GN 101F6  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MILLER N., KRAMER J., ELLIOTT G., KEPPLER D.;  
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA WATERSTON R.;  
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA BADER S., LEE C.-C., LATIF F., SEKIDO Y., DUH F.-M., WEI M.-H.,  
 RA CUNDIFF S., LERMAN M.I., MINNA J.D.;  
 RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 DR EMBL; AC002481; G2340094; -  
 DR EMBL; AF040704; G2935320; -  
 SQ SEQUENCE 222 AA; 23973 MW; 5E50BFD9 CRC32;

Query Match 2.9%; Score 7; DB 4; Length 222;  
 Best Local Similarity 100.0%; Pred. No. 3.41e+00;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 92 LALLCAL 98  
 QY 16 LALLCAL 22  
 |||||

RESULT 12

ID P80943 PRELIMINARY; PRT; 232 AA.  
AC P80943;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DE T-CELL SURFACE GLYCOPROTEIN CD1, CLONE SC0110 (FRAGMENT).  
OS OVIS ARIES (SHEEP).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDEA; BOVIDAE; CAPRINAE; OVIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=FETAL THYMOCYTES;  
RC MEDLINE: 96269982.  
RA FERGUSON E.E., DUTTA B.M., HEIN W.R., HOPKINS J.;  
RT "the sheep CD1 gene family contains at least four CD1B homologues";  
RL IMMUNOGENETICS 44:86-96(1996).  
CC -1- FUNCTION: NOT KNOWN.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- SUBUNIT: ASSOCIATES NON-COVALENTLY WITH BETA-2-MICROGLOBULIN.  
CC -1- SIMILARITY: TO OTHER CD1 ANTIGENS, AND TO MHC CLASS I ANTIGENS.  
CC -1- SIMILARITY: BELONGS TO THE IMMNOGLOBULIN SUPERFAMILY.  
DR EMBL: X90567; G945011; -;  
KW GLYCOPROTEIN; IMMUNOGLOBULIN FOLD; T-CELL.  
FT NON TER 1 1  
FT DOMAIN <1 8 EXTRACELLULAR ALPHA-1.  
FT DOMAIN 9 101 EXTRACELLULAR ALPHA-2.  
FT DOMAIN 102 194 EXTRACELLULAR ALPHA-3.  
FT TRANSMEM 195 223 POTENTIAL.  
FT DOMAIN 224 232 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 45 45 POTENTIAL.  
SQ SEQUENCE 232 AA; 26023 MW; 67BD89CF CRC32;  
Query Match 2.9%; Score 7; DB 7; Length 232;  
Best Local Similarity 100.0%; Pred. No. 3.41e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 115 GPGRLL 121  
QY 35 GPGRLL 41  
RESULT 13  
ID Q31618 PRELIMINARY; PRT; 256 AA.  
AC Q31618;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DE XBT PROTEIN.  
OS BACILLUS SUBTILIS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
OC BACILLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE: 98044033.  
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,  
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,  
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
RA CHOI S.K., CODANI J.J., CUNERTON I.F., CUMMINGS N.J., DANIEL R.A.,  
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,  
RA ENTIAN K.D., ERRINGTON J., FABBET C., FERRARI E., FOULGER D.,  
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,  
RA GIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,  
RA GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,  
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,  
RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,  
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,  
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,  
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,  
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,  
RA NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,

RA PARRO V., POHL T.M., PORTETELLE D., PORKOLLIK S., PRESCOTT A.M.,  
RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,  
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADATE Y.,  
RA SATO T., SCANLAN E., SCHLEICH S., SCHROTER R., SCOFFONE F.,  
RA SERIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,  
RA SOROKIN A., TACCONE E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,  
RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,  
RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZNEGER T., YATA K.,  
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,  
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;  
RT "The complete genome sequence of the gram-positive bacterium Bacillus  
subtilis";  
RL NATURE 390:249-256(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Z99110; E1183189; -;  
SQ SEQUENCE 256 AA; 27022 MW; 01200290 CRC32;  
Query Match 2.9%; Score 7; DB 2; Length 256;  
Best Local Similarity 100.0%; Pred. No. 3.41e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 14 RLLGTG 20  
QY 38 RLLGTG 44  
RESULT 14  
ID Q77265 PRELIMINARY; PRT; 264 AA.  
AC Q77265;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DE EG:22E5.6 PROTEIN.  
GN EG:22E5.6.  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MURPHY L., HARRIS D., BARRELL B.;  
RT "Sequencing the distal X chromosome of Drosophila melanogaster";  
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA BENOS P.;  
RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AL031765; E1329902; -;  
SQ SEQUENCE 264 AA; 29352 MW; 2486C463 CRC32;  
Query Match 2.9%; Score 7; DB 5; Length 264;  
Best Local Similarity 100.0%; Pred. No. 3.41e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 59 ETQLLL 65  
QY 201 ETQLLL 207  
RESULT 15  
ID Q37402 PRELIMINARY; PRT; 332 AA.  
AC Q37402;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DE NADH DEHYDROGENASE, SUBUNIT 1 (EC 1.6.5.3).  
GN NADH.  
OS ALLOMYCES MACROGYNUS.

OG MITOCHONDRION.  
 OC EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES; BLASTOCLADIALES;  
 OC BLASTOCLADIACEAE; ALLAMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96226032.  
 RA PAQUIN B., LANG B.F.;  
 RT "The mitochondrial DNA of Allomyces macrogynus: the complete genomic  
 sequence from an ancestral fungus.";  
 RL J. MOL. BIOL. 255:688-701(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA PAQUIN B., LAFOREST M.J., LANG B.F.;  
 RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: U41288; G1236426;  
 DR PFAM: PF00146; NADHdh; 1.  
 KW OXIDOREDUCTASE; MITOCHONDRION; NAD; TRANSMEMBRANE.  
 SQ SEQUENCE 332 AA; 36515 MW; 0075E871 CRC32;

Query Match 2.9%; Score 7; DB 8; Length 332;  
 Best Local Similarity 100.0%; Pred.No. 3.41e+00;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 25 VLLAVA 31  
 |||||  
 QY 168 VLLAVA 174

Search completed: Mon Jul 12 12:14:47 1999  
 Job time : 27 secs.

97 6 2.5 2035 1 US-08-046- Sequence 5, Applicatio 1.07e+02  
98 5 2.1 112 3 PCT-US95-0 Sequence 4, Applicatio 1.01e+03  
99 5 2.1 213 4 5212074-6 Patent No. 5212074. 1.01e+03  
100 5 2.1 322 4 5212074-1 Patent No. 5212074. 1.01e+03

## ALIGNMENTS

RESULT 1  
ID US-07-725-083-2 STANDARD; PRT: 782 AA.  
XX xxxxxx  
AC  
DT  
XX  
XX  
DE  
XX  
Sequence 2, Application US/07725083  
CC  
CC Sequence 2, Application US/07725083  
CC Patent No. 5407821  
CC GENERAL INFORMATION:  
CC APPLICANT: Breakfield, Xandra O.  
CC APPLICANT: Ozellus, Laurie J.  
CC TITLE OF INVENTION: Genetic Diagnosis of Torsion Dystonia  
CC NUMBER OF SEQUENCES: 2  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
CC STREET: 1100 New York Avenue, Suite 600  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20005  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07725.083  
CC FILING DATE: 03-JUL-1991  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/353,432  
CC FILING DATE: 18-MAY-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Goldstein, Jorge A.  
CC REGISTRATION NUMBER: 29,021  
CC REFERENCE/DOCKET NUMBER: 0609.2170001  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202) 371-2500  
CC TELEFAX: (202) 371-2540  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 782 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 782 AA; 85715 MW; 3053273 CN;

Query Match 3.3%; Score 8; DB 1; Length 782;  
Best Local Similarity 100.0%; Pred. No. 6.65e-01;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 ALLCALS 16  
| | | | |  
Qy 17 ALLCALS 24

RESULT 2  
ID US-08-289-699A-4 STANDARD; PRT: 335 AA.  
XX xxxxxx  
AC  
XX  
DT

XX Sequence 4, Application US/08289699A  
DE Patent No. 5695993  
XX Sequence 4, Application US/08289699A  
CC GENERAL INFORMATION:  
CC APPLICANT: Fukudome, Kenji  
CC APPLICANT: Esmon, Charles T.  
CC TITLE OF INVENTION: Cloning and Regulation of an Endothelial  
CC TITLE OF INVENTION: Cell Protein C/Activated Protein C Receptor  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Patrea L. Pabst  
CC STREET: 2800 One Atlantic Center, 1201 West Peachtree  
CC CITY: Street  
CC STATE: Georgia  
CC COUNTRY: US  
CC ZIP: 30306-3450  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/289,699A  
CC FILING DATE: 12-AUG-1994  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Pabst, Patrea L.  
CC REGISTRATION NUMBER: 31,284  
CC REFERENCE/DOCKET NUMBER: OMRF152  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (404)873-8794  
CC TELEFAX: (404)873-8795  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 335 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC SEQUENCE 335 AA; 37717 MW; 616437 CN;

Query Match 2.9%; Score 7; DB 1; Length 335;  
Best Local Similarity 100.0%; Pred. No. 9.15e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 216 GPGRLLL 222  
| | | | |  
Qy 35 GPGRLLL 41

RESULT 3  
ID US-08-878-283-4 STANDARD; PRT: 335 AA.  
XX xxxxxx  
AC  
XX  
DT  
XX

Sequence 4, Application US/08878283  
CC Patent No. 5852171  
CC GENERAL INFORMATION:  
CC APPLICANT: Fukudome, Kenji  
CC APPLICANT: Esmon, Charles T.  
CC TITLE OF INVENTION: Cloning and Regulation of an Endothelial  
CC TITLE OF INVENTION: Cell Protein C/Activated Protein C Receptor  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Patrea L. Pabst  
CC STREET: 2800 One Atlantic Center, 1201 West Peachtree

\*\*\*\*\*

WISORLH (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
on: Mon Jul 12 12:15:07 1999; MasPar time 4.44 Seconds  
551.153 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-911-423-4  
Description: (1-241) from US08911423.pep  
Perfect Score: 241  
Sequence: 1 MAQHGAAGAFRALCGLALLC.....EEERGSERAEKGRGLDIW 241

Scoring table: . TABLE unitprotatable  
Gap 60

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 08  
Listing first 100 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 2.645; Variance 0.703; scale 3.762

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	Query Match	ID	Description	Pred. No.
1	8	3.3	782	US-07-725- Sequence 2, Applicatio	6.65e-01
2	7	2.9	335	US-08-289- Sequence 4, Applicatio	9.15e+00
3	7	2.9	335	US-08-878- Sequence 4, Applicatio	9.15e+00
4	7	2.9	713	US-08-332- Sequence 56, Applicati	9.15e+00
5	7	2.9	713	US-08-188- Sequence 62, Applicati	9.15e+00
6	7	2.9	713	US-08-332- Sequence 62, Applicatio	9.15e+00
7	7	2.9	1218	US-08-400- Sequence 6, Applicatio	9.15e+00
8	6	2.5	13	PCT-US92-0 Sequence 2, Applicatio	1.07e+02
9	6	2.5	15	PCT-US93-0 Sequence 23, Applicatio	1.07e+02
10	6	2.5	29	US-08-376- Sequence 18, Applicati	1.07e+02
11	6	2.5	34	US-08-190- Sequence 165, Applicat	1.07e+02
12	6	2.5	79	4 5284931-7 Patent No. 5284931.	1.07e+02
13	6	2.5	93	PCT-US95-0 Sequence 2, Applicatio	1.07e+02
14	6	2.5	93	US-08-480- Sequence 2, Applicatio	1.07e+02
15	6	2.5	163	PCT-US95-0 Sequence 11, Applicati	1.07e+02
16	6	2.5	163	US-08-469- Sequence 11, Applicati	1.07e+02
17	6	2.5	179	US-08-665- Sequence 8, Applicatio	1.07e+02
18	6	2.5	185	4 5489533-4 Patent No. 5489533.	1.07e+02
19	6	2.5	202	1 US-08-048- Sequence 2, Applicatio	1.07e+02
20	6	2.5	223	2 US-08-928- Sequence 1, Applicatio	1.07e+02
21	6	2.5	277	1 US-08-400- Sequence 1, Applicatio	1.07e+02
22	6	2.5	297	2 US-08-478- Sequence 6, Applicatio	1.07e+02
23	6	2.5	297	1 US-08-077- Sequence 6, Applicatio	1.07e+02

24	6	2.5	297	1 US-07-866- Sequence 6, Applicatio	1.07e+02
25	6	2.5	324	2 US-08-816- Sequence 2, Applicatio	1.07e+02
26	6	2.5	331	1 US-08-136- Sequence 1, Applicatio	1.07e+02
27	6	2.5	341	1 US-08-190- Sequence 45, Applicati	1.07e+02
28	6	2.5	363	2 US-08-784- Sequence 15, Applicati	1.07e+02
29	6	2.5	367	2 US-08-132- Sequence 2, Applicatio	1.07e+02
30	6	2.5	367	3 PCT-US92-0 Sequence 2, Applicatio	1.07e+02
31	6	2.5	379	1 US-08-887- Sequence 36, Applicati	1.07e+02
32	6	2.5	406	1 US-08-136- Sequence 13, Applicati	1.07e+02
33	6	2.5	446	2 US-08-665- Sequence 10, Applicati	1.07e+02
34	6	2.5	453	4 5284931-3 Patent No. 5284931.	1.07e+02
35	6	2.5	453	2 US-08-132- Sequence 6, Applicatio	1.07e+02
36	6	2.5	453	3 PCT-US92-0 Sequence 6, Applicatio	1.07e+02
37	6	2.5	480	2 US-08-425- Sequence 11, Applicati	1.07e+02
38	6	2.5	507	2 US-08-469- Sequence 14, Applicati	1.07e+02
39	6	2.5	507	1 US-08-464- Sequence 14, Applicati	1.07e+02
40	6	2.5	509	1 US-08-445- Sequence 8, Applicatio	1.07e+02
41	6	2.5	531	3 PCT-US95-0 Sequence 3, Applicatio	1.07e+02
42	6	2.5	531	2 US-08-789- Sequence 3, Applicatio	1.07e+02
43	6	2.5	531	2 US-08-752- Sequence 3, Applicatio	1.07e+02
44	6	2.5	532	4 5284931-2 Patent No. 5284931.	1.07e+02
45	6	2.5	532	2 US-08-483- Sequence 118, Applicat	1.07e+02
46	6	2.5	532	1 US-07-618- Sequence 1, Applicatio	1.07e+02
47	6	2.5	532	1 US-08-196- Sequence 3, Applicatio	1.07e+02
48	6	2.5	532	2 US-08-689- Sequence 12, Applicati	1.07e+02
49	6	2.5	533	2 US-08-458- Sequence 10, Applicati	1.07e+02
50	6	2.5	533	1 US-08-368- Sequence 10, Applicati	1.07e+02
51	6	2.5	533	3 PCT-US93-0 Sequence 10, Applicati	1.07e+02
52	6	2.5	547	2 US-08-482- Sequence 1, Applicatio	1.07e+02
53	6	2.5	547	1 US-08-473- Sequence 6, Applicatio	1.07e+02
54	6	2.5	547	1 US-08-314- Sequence 1, Applicatio	1.07e+02
55	6	2.5	547	1 US-08-473- Sequence 1, Applicatio	1.07e+02
56	6	2.5	547	1 US-08-314- Sequence 1, Applicatio	1.07e+02
57	6	2.5	547	1 US-08-433- Sequence 1, Applicatio	1.07e+02
58	6	2.5	547	2 US-08-487- Sequence 1, Applicatio	1.07e+02
59	6	2.5	547	2 US-08-483- Sequence 1, Applicatio	1.07e+02
60	6	2.5	547	2 US-08-483- Sequence 1, Applicatio	1.07e+02
61	6	2.5	622	2 PCT-US92-0 Sequence 4, Applicatio	1.07e+02
62	6	2.5	622	2 US-08-132- Sequence 6, Applicatio	1.07e+02
63	6	2.5	627	2 US-08-466- Sequence 6, Applicatio	1.07e+02
64	6	2.5	629	1 US-08-278- Sequence 6, Applicatio	1.07e+02
65	6	2.5	629	2 US-08-132- Sequence 8, Applicatio	1.07e+02
66	6	2.5	629	3 PCT-US93-0 Sequence 8, Applicatio	1.07e+02
67	6	2.5	644	3 PCT-US93-0 Sequence 21, Applicati	1.07e+02
68	6	2.5	668	2 US-08-786- Sequence 13, Applicati	1.07e+02
69	6	2.5	668	1 US-08-232- Sequence 13, Applicati	1.07e+02
70	6	2.5	676	1 US-08-343- Sequence 2, Applicatio	1.07e+02
71	6	2.5	676	1 US-08-343- Sequence 8, Applicatio	1.07e+02
72	6	2.5	676	2 US-08-462- Sequence 2, Applicatio	1.07e+02
73	6	2.5	676	2 US-08-462- Sequence 2, Applicatio	1.07e+02
74	6	2.5	736	3 PCT-US93-0 Sequence 15, Applicati	1.07e+02
75	6	2.5	740	3 PCT-US93-0 Sequence 17, Applicati	1.07e+02
76	6	2.5	780	1 US-08-188- Sequence 50, Applicati	1.07e+02
77	6	2.5	780	1 US-08-332- Sequence 44, Applicati	1.07e+02
78	6	2.5	780	2 US-08-332- Sequence 50, Applicati	1.07e+02
79	6	2.5	788	2 US-08-786- Sequence 15, Applicati	1.07e+02
80	6	2.5	788	1 US-08-232- Sequence 15, Applicati	1.07e+02
81	6	2.5	920	2 US-08-451- Sequence 2, Applicatio	1.07e+02
82	6	2.5	937	1 US-08-253- Sequence 31, Applicati	1.07e+02
83	6	2.5	1005	1 US-08-478- Sequence 3, Applicatio	1.07e+02
84	6	2.5	1005	2 US-08-717- Sequence 3, Applicatio	1.07e+02
85	6	2.5	1005	1 US-08-089- Sequence 3, Applicatio	1.07e+02
86	6	2.5	1005	2 US-08-266- Sequence 3, Applicatio	1.07e+02
87	6	2.5	1005	3 PCT-US94-0 Sequence 3, Applicatio	1.07e+02
88	6	2.5	1052	2 US-08-852- Sequence 2, Applicatio	1.07e+02
89	6	2.5	1196	1 US-08-144- Sequence 4, Applicatio	1.07e+02
90	6	2.5	1356	2 US-07-930- Sequence 8, Applicatio	1.07e+02
91	6	2.5	1356	2 US-08-810- Sequence 8, Applicatio	1.07e+02
92	6	2.5	1528	2 US-08-463- Sequence 6, Applicatio	1.07e+02
93	6	2.5	1528	2 US-08-462- Sequence 6, Applicatio	1.07e+02
94	6	2.5	1958	1 US-07-945- Sequence 2, Applicatio	1.07e+02
95	6	2.5	2035	3 PCT-US93-1 Sequence 5, Applicatio	1.07e+02
96	6	2.5	2035	1 US-08-393- Sequence 5, Applicatio	1.07e+02

Query Match 2.5%; Score 6; DB 3; Length 163;  
Best Local Similarity 100.0%; Pred. No. 1.07e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 4 LALLCA 9  
| | | | |  
QY 16 LALLCA 21

Search completed: Mon Jul 12 12:15:16 1999  
Job time : 9 secs.

CC ADDRESSEE: CECCHI, STEWART & OLSTEIN  
CC STREET: 6 BECKER FARM ROAD  
CC CITY: ROSELAND  
CC STATE: NEW JERSEY  
CC COUNTRY: USA  
CC ZIP: 07068  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: 3.5 INCH DISKETTE  
CC OPERATING SYSTEM: MS-DOS  
CC SOFTWARE: WORD PERFECT 5.1  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US95/07294  
CC FILING DATE: June 6, 1995  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/464,594  
CC FILING DATE: June 5, 1995  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: FERRARO, GREGORY D.  
CC REGISTRATION NUMBER: 36,134  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 201-994-1700  
CC TELEFAX: 201-994-1744  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 93 AMINO ACIDS  
CC TYPE: AMINO ACID  
CC STRANDEDNESS:  
CC TOPOLOGY: LINEAR  
CC MOLECULE TYPE: PROTEIN  
CC SEQUENCE 93 AA; 10580 MW; 48196 CN;  
  
Query Match 2.5%; Score 6; DB 3; Length 93;  
Best Local Similarity 100.0%; Pred. No. 1.07e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 13 VLLAVA 18  
QY 169 VLLAVA 174  
  
RESULT 14  
ID US-08-480-449-2 STANDARD; PRT; 93 AA.  
XX  
XX  
Sequence 2, Application US/08480449  
Patent No. 5688927  
GENERAL INFORMATION:  
APPLICANT: Godiska, Ronald  
APPLICANT: Gray, Patrick W.  
TITLE OF INVENTION: MACROPHAGE DERIVED CHEMOKINE  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,449

CC FILING DATE:  
CC CLASSIFICATION: 530  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Gass, David A.  
CC REGISTRATION NUMBER: 38,153  
CC REFERENCE/DOCKET NUMBER: 27866/32779  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 312/474-6300  
CC TELEFAX: 312/474-0448  
CC TELEX: 25-3856  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 93 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 93 AA; 10580 MW; 48196 CN;  
  
Query Match 2.5%; Score 6; DB 1; Length 93;  
Best Local Similarity 100.0%; Pred. No. 1.07e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 13 VLLAVA 18  
QY 169 VLLAVA 174  
  
RESULT 15  
ID PCT-US95-07289-11 STANDARD; PRT; 163 AA.  
XX  
XX  
Sequence 11, Application PC/TUS9507289  
Sequence 11, Application PC/TUS9507289  
GENERAL INFORMATION:  
APPLICANT: Yu, Guo-Liang  
APPLICANT: Rosen, Craig  
TITLE OF INVENTION: Colon Specific Genes and Proteins  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
ADDRESSEE: Stewart & Olstein  
STREET: 6 Becker Farm Road  
CITY: Roseland  
STATE: NJ  
COUNTRY: USA  
ZIP: 07068-1739  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07289  
FILING DATE: 06-JUN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferraro, Gregory D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-265  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 163 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 163 AA; 17724 MW; 152889 CN;

CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/376,362A  
CC FILING DATE: 23-JAN-1995  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Kegan A., Sarah  
CC REGISTRATION NUMBER: 32,141  
CC REFERENCE/DOCKET NUMBER: 01107.48125  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202 508-9100  
CC TELEFAX: 202-508-9299  
CC INFORMATION FOR SEQ ID NO: 18:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 29 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC HYPOTHETICAL: YES  
CC ANTI-SENSE: NO  
CC FRAGMENT TYPE: internal  
CC ORIGINAL SOURCE:  
CC ORGANISM: Rattus rattus  
CC TISSUE TYPE: taste tissue  
CC  
SQ SEQUENCE 29 AA; 3282 MW; 4545 CN;  
  
Query Match 2.5%; Score 6; DB 1; Length 29;  
Best Local Similarity 100.0%; Pred. No. 1.07e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 12 GRLLG 17  
| | | | |  
QY 37 GRLLG 42  
  
RESULT 11  
ID US-08-190-802A-165 STANDARD; PRT; 34 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
XX  
SQ SEQUENCE 165, Application US/08190802A  
Patent No. 5519003  
GENERAL INFORMATION:  
CC APPLICANT: Mochly-Rosen, Daria  
CC APPLICANT: Ron, Dorit  
CC TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
CC TITLE OF INVENTION: Thereof  
CC NUMBER OF SEQUENCES: 265  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Dehlinger & Associates  
CC STREET: P.O. Box 60850  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94306-0850  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/190,802A  
CC FILING DATE: 01-FEB-1994  
CC CLASSIFICATION: 530  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Fabian, Gary R.

CC REGISTRATION NUMBER: 33,875  
CC REFERENCE/DOCKET NUMBER: 8600-0139  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 324-0880  
CC TELEFAX: (415) 324-0960  
CC INFORMATION FOR SEQ ID NO: 165:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 34 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: peptide  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC ORIGINAL SOURCE:  
CC INDIVIDUAL ISOLATE: GTP binding prt squid rv, Fig. 28  
SQ SEQUENCE 34 AA; 3718 MW; 5886 CN;  
  
Query Match 2.5%; Score 6; DB 1; Length 34;  
Best Local Similarity 100.0%; Pred. No. 1.07e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 18 GRLLG 23  
| | | | |  
QY 37 GRLLG 42  
  
RESULT 12  
ID 5284931-7 STANDARD; PRT; 85 AA.  
XX  
AC xxxxxx  
XX  
DT 01-JAN-1900  
XX  
DE  
XX  
XX Patent No. 5284931.  
CC Patent No. 5284931  
CC APPLICANT: SPRINGER, TIMOTHY A.; ROTHLEIN, ROBERT; MARLIN,  
CC STEVEN D.; DUSTIN, MICHAEL L.  
CC TITLE OF INVENTION: INTERCELLULAR ADHESION MOLECULES AND  
CC THEIR BINDING LIGANDS  
CC NUMBER OF SEQUENCES: 41  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/515,478  
CC FILING DATE: 27-APR-1990  
CC SEQ ID NO: 7  
CC LENGTH: 79  
CC SEQUENCE 85 AA; 9355 MW; 38426 CN;  
SQ  
  
Query Match 2.5%; Score 6; DB 4; Length 79;  
Best Local Similarity 100.0%; Pred. No. 1.07e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 31 LTVVLL 36  
| | | | |  
QY 166 LTVVLL 171  
  
RESULT 13  
ID PCT-US95-07294-2 STANDARD; PRT; 93 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
XX  
SQ SEQUENCE 2, Application PC/TUS9507294  
Sequence 2, Application PC/TUS9507294  
GENERAL INFORMATION:  
CC APPLICANT: LI, ET AL.  
CC TITLE OF INVENTION: Human Chemokine Beta-13  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CC

RESULT 8  
ID PCT-US92-01196-2 STANDARD; PRT; 13 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
XX  
XX Sequence 2, Application PC/TUS9201196  
XX  
XX Sequence 2, Application PC/TUS9201196  
CC GENERAL INFORMATION:  
CC APPLICANT: Shimasaki, Shunichi  
CC APPLICANT: Ling, Nicholas C.  
CC TITLE OF INVENTION: Insulin-Like Growth Factor Binding  
CC TITLE OF INVENTION: Protein  
CC NUMBER OF SEQUENCES: 10  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Fitch, Even, Tabin & Flannery  
CC STREET: 135 South LaSalle Street, Suite 900  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: United States  
CC ZIP: 60603-4277  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US92/01196  
CC FILING DATE: 19920213  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/658,410  
CC FILING DATE: 14-FEB-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Watt, Phillip H.  
CC REGISTRATION NUMBER: 25,939  
CC REFERENCE/DOCKET NUMBER: 51145PCT  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312)372-7842  
CC TELEFAX: (312)372-7848  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 13 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC SEQUENCE 13 AA; 1143 MW; 1131 CN;  
  
Query Match 2.5%; Score 6; DB 3; Length 13;  
Best Local Similarity 75.0%; Pred. No. 1.07e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Db 4 GXPGXGPG 11  
QY 30 GPGCGPG 37  
  
RESULT 9  
ID PCT-US93-06751-23 STANDARD; PRT; 15 AA.  
XX  
AC xxxxxx  
XX  
XX  
XX  
XX Sequence 23, Application PC/TUS9306751  
XX  
XX Sequence 23, Application PC/TUS9306751  
CC GENERAL INFORMATION:  
CC APPLICANT: P. Keller, A. J. Conley, A. R. Shaw, B. A. Arnold  
CC TITLE OF INVENTION: Immunological Conjugates of OMPC and  
CC TITLE OF INVENTION: HIV-Specific Selected Principal Neutralization GXG Epitope

CC NUMBER OF SEQUENCES: 146  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Merck & Co., Inc.  
CC STREET: P. O. Box 2000  
CC CITY: Rahway  
CC STATE: NJ  
CC COUNTRY: USA  
CC ZIP: 07065  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/06751  
CC FILING DATE: 19930719  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Meredith, Roy D.  
CC REGISTRATION NUMBER: 30,777  
CC REFERENCE/DOCKET NUMBER: 18614  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (908) 594-4678  
CC TELEFAX: (908) 594-4720  
CC TELEX: 138825  
CC INFORMATION FOR SEQ ID NO: 23:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 15 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC HYPOTHETICAL: NO  
CC ANTI-SENSE: NO  
CC IMMEDIATE SOURCE: Random Epitope Library Alpha  
CC SEQUENCE 15 AA; 1677 MW; 1247 CN;  
  
Query Match 2.5%; Score 6; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.07e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 8 GPGRL 13  
QY 35 GPGRL 40  
  
RESULT 10  
ID US-08-376-362A-18 STANDARD; PRT; 29 AA.  
XX  
AC xxxxxx  
XX  
XX  
XX  
XX  
XX Sequence 18, Application US/08376362A  
XX  
XX Sequence 18, Application US/08376362A  
CC Patent No. 5693756  
CC GENERAL INFORMATION:  
CC APPLICANT: Li, Xiao-Jiang  
CC APPLICANT: Blackshaw, Seth  
CC APPLICANT: Snyder, Solomon H.  
CC TITLE OF INVENTION: AMILORIDE-SENSITIVE SODIUM CHANNEL AND  
CC TITLE OF INVENTION: METHOD OF IDENTIFYING SUBSTANCES WHICH STIMULATE OR BLO  
CC TITLE OF INVENTION: SALTY TASTE PERCEPTION  
CC NUMBER OF SEQUENCES: 20  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Banner & Allegretti, LTD  
CC STREET: 1001 G Street, N.W., Eleventh Floor  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: U.S.A.  
CC ZIP: 20001-4597  
CC COMPUTER READABLE FORM:

CC TELEPHONE: (312) 474-6300  
CC TELEFAX: (312) 474-0448  
CC TELEX: 25-3856  
CC INFORMATION FOR SEQ ID NO: 62:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 713 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 713 AA; 78286 MW; 2721372 CN;

Query Match 2.9%; Score 7; DB 1; Length 713;  
Best Local Similarity 100.0%; Pred. No. 9.15e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 VLLTSA 22  
QY 177 VLLTSA 183

ULT 6  
US-08-332-638-62 STANDARD; PRT; 713 AA.  
xxxxxx

Sequence 62, Application US/08332638

Sequence 62, Application US/08332638  
Patent No. 5646250

GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
ADDRESSEE: Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/332,638  
FILING DATE: 01-NOV-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/872,643  
FILING DATE: 17 APR 1992  
APPLICATION NUMBER: US/08/049,460  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5646250 and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31340  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 713 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 713 AA; 78286 MW; 2721372 CN;

Query Match 2.9%; Score 7; DB 1; Length 713;  
Best Local Similarity 100.0%; Pred. No. 9.15e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 VLLTSA 22  
QY 177 VLLTSA 183

Query Match 2.9%; Score 7; DB 1; Length 713;  
Best Local Similarity 100.0%; Pred. No. 9.15e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 VLLTSA 22  
QY 177 VLLTSA 183

RESULT 7  
ID US-08-400-159-6 STANDARD; PRT; 1218 AA.  
XX  
AC xxxxxx  
XX

Sequence 6, Application US/08400159

Sequence 6, Application US/08400159  
Patent No. 5869282

GENERAL INFORMATION:  
APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myat, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
APPLICANT: Gray, Grace E.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/400,159  
FILING DATE: 07-MAR-1995  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1218 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 1218 AA; 133810 MW; 7064054 CN;

Query Match 2.9%; Score 7; DB 2; Length 1218;  
Best Local Similarity 100.0%; Pred. No. 9.15e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 18 LALLCAL 24  
QY 16 LALLCAL 22

CC STREET: Street  
CC CITY: Atlanta  
CC STATE: Georgia  
CC COUNTRY: US  
CC ZIP: 30306-3450  
CC COMPUTER READABLE FORM: disk  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/878,283  
CC FILING DATE:  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/289,699  
CC FILING DATE: 12-AUG-1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Padst, Patrea L.  
CC REGISTRATION NUMBER: 31,284  
CC REFERENCE/DOCKET NUMBER: OMRE152  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (404)873-8794  
CC TELEFAX: (404)873-8795  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 335 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC SEQUENCE 335 AA; 37717 MW; 616437 CN;  
  
Query Match 2.9%; Score 7; DB 2; Length 335;  
Best Local Similarity 100.0%; Pred. No. 9.15e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 216 GPGRLLL 222  
|||||  
QY 35 GPGRLLL 41  
  
RESULT 4  
ID US-08-332-643-56 STANDARD; PRT; 713 AA.  
XX XXXXX  
AC  
XX  
XX  
  
Sequence 56, Application US/08332643  
XX  
CC Sequence 56, Application US/08332643  
CC Patent No. 5639634  
CC GENERAL INFORMATION:  
CC APPLICANT: Suzuki, Shintaro  
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
CC NUMBER OF SEQUENCES: 56  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
CC STREET: Two First National Plaza, 20 South Clark  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60603  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/332,643

CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/872,643  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5639634and, Greta E.  
CC REGISTRATION NUMBER: 35,302  
CC REFERENCE/DOCKET NUMBER: 27866/30795  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (312) 346-5750  
CC TELEFAX: (312) 984-9740  
CC TELEX: 25-3856  
CC INFORMATION FOR SEQ ID NO: 56:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 713 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 713 AA; 78286 MW; 2721372 CN;  
  
Query Match 2.9%; Score 7; DB 1; Length 713;  
Best Local Similarity 100.0%; Pred. No. 9.15e+00;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 16 VLLLTSA 22  
|||||  
QY 177 VLLLTSA 183  
  
RESULT 5  
ID US-08-188-228-62 STANDARD; PRT; 713 AA.  
XX XXXXX  
AC  
XX  
DT  
XX  
DE Sequence 62, Application US/08188228  
XX  
CC Sequence 62, Application US/08188228  
CC Patent No. 5597725  
CC GENERAL INFORMATION:  
CC APPLICANT: Suzuki, Shintaro  
CC TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
CC NUMBER OF SEQUENCES: 62  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
CC ADDRESSEE: Borun  
CC STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CC CITY: Chicago  
CC STATE: Illinois  
CC COUNTRY: USA  
CC ZIP: 60606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/188,228  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/049,460  
CC FILING DATE: 19 APR 1993  
CC APPLICATION NUMBER: US 07/872,643  
CC FILING DATE: 17 APR 1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: No. 5597725and, Greta E.  
CC REGISTRATION NUMBER: 35,302  
CC REFERENCE/DOCKET NUMBER: 31340  
CC TELECOMMUNICATION INFORMATION:

\*\*\*\*\*

WQSEFL

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
On: Mon Jul 12 12:08:11 1999; MasPar time 15.41 Seconds  
Tabular output not generated. 314.729 Million cell updates/sec

Title: >US-08-911-423-2  
Description: (1-228) from US08911423.ppt  
Perfect Score: 228  
Sequence: 1 MGAWMLYGVSMCLVLDLQ.....PEEERGETEKEKHLGGRWP 228

Scoring table: TABLE unitprotable  
Gap 60  
Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 2.686; Variance 0.622; scale 4.320

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

length  
% of match

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	228	100.0	228	31	W37838 Amino acid sequence o	0.00e+00
2	228	100.0	228	33	W49016 Mouse glucocorticoid	0.00e+00
3	188	82.5	222	33	W49018 Mouse glucocorticoid	0.00e+00
4	188	82.5	224	33	W49017 Mouse glucocorticoid	0.00e+00
5	11	4.8	228	31	W37840 Truncated human 312C2	1.99e-05
6	11	4.8	241	31	W37839 Amino acid sequence o	1.99e-05
7	11	4.8	311	31	W37842 Human 312C2 protein f	1.99e-05
8	6	2.6	69	39	W67870 Human secreted protei	9.87e+01
9	6	2.6	140	29	W55491 H. pylori ORF hp2p109	9.87e+01
10	6	2.6	165	29	W55326 H. pylori ORF hp2p106	9.87e+01
11	6	2.6	167	6	R29897 HCV NS4-NS5 peptide N	9.87e+01
12	6	2.6	171	6	R29896 HCV NS4-NS5 peptide N	9.87e+01
13	6	2.6	171	1	P90160 Sequence of hepatitis	9.87e+01
14	6	2.6	171	1	P92043 Sequence encoded in t	9.87e+01
15	6	2.6	171	6	R29874 HCV NS4-NS5 peptide N	9.87e+01
16	6	2.6	171	6	R29895 HCV NS4-NS5 peptide N	9.87e+01

Human 312C2 protein f 9.87e+01  
HCV polyptide 16. 9.87e+01  
Hepatitis C virus ant 9.87e+01  
HCV NS4-NS5 peptide 2 9.87e+01  
Human HMBU14 protein 9.87e+01  
STRL33 protein sequen 9.87e+01  
Feline herpes virus t 9.87e+01  
tvb polyptide. 9.87e+01  
Recombinant protein p 9.87e+01  
HCV CKS-NS5G - pHCV-5 9.87e+01  
HCV CKS-NS5G recombin 9.87e+01  
Hepatitis C virus ant 9.87e+01  
Hepatitis C virus ant 9.87e+01  
Hepatitis C virus ant 9.87e+01  
Antigenic portion of 9.87e+01  
Hepatitis C virus opt 9.87e+01  
NS5B protein (residue 9.87e+01  
Recombinant modified 9.87e+01  
Hepatitis C virus RNA 9.87e+01  
HCV NS4-NS5 peptide 1 9.87e+01  
Recombinant modified 9.87e+01  
Modified HCV RNA-depe 9.87e+01  
Modified HCV RNA-depe 9.87e+01  
Modified HCV RNA-depe 9.87e+01  
Fragment of 101 KD pr 9.87e+01  
Gene 036 product diff 9.87e+01  
Modified retinoblasto 9.87e+01  
Retinoblastoma 94KD t 9.87e+01  
Human retinoblastoma 9.87e+01  
Protein sequence of t 9.87e+01  
Modified retinoblasto 9.87e+01  
Modified retinoblasto 9.87e+01  
HCV NS4-NS5 peptide 2 9.87e+01  
p100 protein from hum 9.87e+01  
Modified retinoblasto 9.87e+01  
Modified retinoblasto 9.87e+01  
Modified retinoblasto 9.87e+01  
Retinoblastoma suscep 9.87e+01  
Human retinoblastoma 9.87e+01  
Dogfish shark kidney 9.87e+01  
Type B human platelet 9.87e+01  
Human platelet-derive 9.87e+01  
Platelet-derived grow 9.87e+01  
PT-NANBH NS5-NS3-core 9.87e+01  
Portion of PT-NANBH v 9.87e+01  
Bacillus subtilis srf 9.87e+01  
HCV NS4-NS5 peptide 1 9.87e+01  
Neuronal nitrogen mon 9.87e+01  
HCV NS2-NS5B non-stru 9.87e+01  
Sequence of viral L43 9.87e+01  
Attenuated hepatitis 9.87e+01  
Hepatitis A virus HM- 9.87e+01  
Sequence of the alpha 9.87e+01  
Human neuronal calciu 9.87e+01  
Human calcium channel 9.87e+01  
Peptide encoded by co 9.87e+01  
Sequence of the alpha 9.87e+01  
Human neuronal calciu 9.87e+01  
Human calcium channel 9.87e+01  
Sequence encoded in t 9.87e+01  
HCV-1 polyprotein. 9.87e+01  
Composite HCV HC-J1/C 9.87e+01  
Composite hepatitis C 9.87e+01  
Blood transmissible N 9.87e+01  
Non-A, non-B viral ge 9.87e+01

12/7/96

90 6 2.6 3010 4 R20111 Non-A, non-B viral ge 9.87e+01  
91 6 2.6 3010 18 R82694 Partial HCV non-struct 9.87e+01  
92 6 2.6 3010 14 R68644 Hepatitis C virus RNA 9.87e+01  
93 6 2.6 3010 13 R68622 HCV protein cleavable 9.87e+01  
94 6 2.6 3011 36 R77398 Hepatitis C virus-H C 9.87e+01  
95 6 2.6 3011 4 R21519 Compiled HCV sequence 9.87e+01  
96 6 2.6 3011 7 R34468 Encoded by full-length 9.87e+01  
97 6 2.6 3011 6 R31621 Hepatitis C virus (HC 9.87e+01  
98 6 2.6 3011 28 W40038 HCV polyprotein. 9.87e+01  
99 6 2.6 3011 26 W34480 HCV polyprotein. 9.87e+01  
100 6 2.6 3014 7 R35207 Hepatitis C virus pro 9.87e+01

## ALIGNMENTS

## RESULT 1

ID W37838 standard; Protein; 228 AA.  
AC W37838;  
DT 28-JUL-1998 (first entry)  
DE Amino acid sequence of the mouse 312C2 T cell protein.  
KW Mouse 312C2 T cell protein; thymus cell; spleen cell; T cell;  
antigen-specific T cell proliferation; cytokine production by T-cell;  
apoptosis; cancer; haematopoietic cells; lymphoid cell;  
autoimmune disorders.  
KW Mus sp.  
OS WO9806842-A1.  
PN 19-FEB-1998.  
PF 14-AUG-1997; U13931.  
PR 07-OCT-1996; US-027901.  
PR 16-AUG-1996; US-689943.  
PA (SCHE ) SCHERING CORP.  
PI Gorman DM, Randall TD, Zlotnik A;  
DR WPI: 98-159534/14.  
DR N-PSDB: V19152.  
PT Isolated 312C2 T cell gene - used to develop products for treating,  
e.g. cancers, auto-immune disorders, transplantation rejection and  
other T cell disorders  
PS Claim 2: Pages 57-58; 71pp; English.  
CC This is the amino acid sequence of the mouse 312C2 T cell protein.  
CC The 312C2 proteins are expressed in thymus cells and are induced on  
T cells and spleen cells following activation. Engagement of 312C2  
stimulates proliferation of T cell clones, antigen-specific  
proliferation and cytokine production by T-cells, and potentiates T  
cell expansion or apoptosis. The products can be used in the  
treatment of conditions associated with abnormal physiology or  
development, including abnormal proliferation, e.g. cancerous  
conditions or degenerative conditions. They can be used in the  
regulation or development of haematopoietic cells, e.g. lymphoid cells  
which affect immunological responses, e.g. autoimmune disorders.  
Sequence 228 AA;

Query Match 100.0%; Score 228; DB 31; Length 228;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mgawamlygvsmclvldlqqpsvveepcgpgkvngsgnntccsllyapgedcpkerc 60  
QY 1 MGAWAMLYGVSMCLVLDLGGPSVVEEPGCGPGKVQNGSGNNTCCSLYAPGEDCPKERC 60  
Db 61 icvtpeyhcgdppqckickhyppcqgrvesqgdvifgfrvacamtgsagrdghcrlwt 120  
QY 61 ICVTPEYHCGDPPQCKICKHYPCQGRVESQGDIVFGFRVCAMGTFSAGRDGHCLWT 120  
Db 121 ncsqfgfltmfpgnknthnavcipeplteqyghltviflvmaaciffittvqlghlwl 180  
QY 121 NCSQFGFLTMFPNGNKNTHNAVCIPEPLTEQYGHLTVIFLVMAACIFFITTVQLGLHIWL 180  
Db 181 rrqhmcpretqpfavqlsaedacsfgfeeergqteekchlgrwp 228  
QY 181 RRQHMCPRETQPFQAEVQLSAEDACSFQFPEERQTEEKCHLGRWP 228

## RESULT 2

ID W49016 standard; Protein; 228 AA.  
AC W49016;  
DT 29-SEP-1998 (first entry)  
DE Mouse glucocorticoid induced TNFR-family related protein (GTR).  
KW Mouse glucocorticoid induced TNFR-family related protein; lymphocyte;  
GTR; tumour necrosis factor; apoptosis; hodgkin's disease; GTR-B;  
GTR-C.  
OS Mus sp.  
EH Key Location/Qualifiers  
FT Peptide 1..19 /note= "Signal peptide"  
FT Region 29..60 /note= "Cysteine pseudorepeat"  
FT Modified\_site 36 /note= "N-glycosylated"  
FT Modified\_site 40 /note= "N-glycosylated"  
FT Region 62..100 /note= "Cysteine pseudorepeat"  
FT Region 103..141 /note= "Cysteine pseudorepeat"  
FT Modified\_site 121 /note= "Cysteine pseudorepeat"  
FT Modified\_site 134 /note= "N-glycosylated"  
FT Region 154..176 /note= "N-glycosylated"  
FT Modified\_site 199 /note= "Probable transmembrane domain"  
FT /note= "Possibly phosphorylated"  
PN WO9824895-A1.  
PD 11-JUN-1998.  
PF 08-NOV-1997; E06252.  
PR 02-DEC-1996; GB-025074.  
PA (PHAA ) PHARMACIA & UPJOHN SPA.  
PI Riccardi C;  
DR WPI: 98-333315/29.  
DR N-PSDB: V32773.  
PT New isolated glucocorticoid induced TNFR related polypeptide - used  
to stimulate lymphocyte activity and cell death rescue, useful to,  
e.g. develop products to suppress lymphocyte activity and induce  
apoptosis  
PS Claim 13; Pages 36-37; 53pp; English.  
CC The present claimed sequence represents a mouse glucocorticoid  
induced TNFR-family related protein (GTR). The invention also  
claims for GTR-B (W49017) and GTR-C (W49018) which are splicing  
variants of GTR. The GTRs are claimed to be useful for stimulating  
lymphocyte activity and cell death rescue. GTR antagonists are  
claimed to be useful for suppressing the lymphocyte activity and for  
inducing apoptotic deletion. GTR cDNAs and the proteins they encode  
are also claimed to be useful for suppressing growth of tumour cells  
over-expressing GTR or for the treatment of refractory hodgkin's  
disease.  
Sequence 228 AA;

Query Match 100.0%; Score 228; DB 33; Length 228;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mgawamlygvsmclvldlqqpsvveepcgpgkvngsgnntccsllyapgedcpkerc 60  
QY 1 MGAWAMLYGVSMCLVLDLGGPSVVEEPGCGPGKVQNGSGNNTCCSLYAPGEDCPKERC 60  
Db 61 icvtpeyhcgdppqckickhyppcqgrvesqgdvifgfrvacamtgsagrdghcrlwt 120  
QY 61 ICVTPEYHCGDPPQCKICKHYPCQGRVESQGDIVFGFRVCAMGTFSAGRDGHCLWT 120  
Db 121 ncsqfgfltmfpgnknthnavcipeplteqyghltviflvmaaciffittvqlghlwl 180  
QY 121 NCSQFGFLTMFPNGNKNTHNAVCIPEPLTEQYGHLTVIFLVMAACIFFITTVQLGLHIWL 180  
Db 181 rrqhmcpretqpfavqlsaedacsfgfeeergqteekchlgrwp 228  
QY 181 RRQHMCPRETQPFQAEVQLSAEDACSFQFPEERQTEEKCHLGRWP 228

PT New isolated glucocorticoid induced TNFR related polypeptide - used  
PT to stimulate lymphocyte activity and cell death rescue, useful to,  
PT e.g. develop products to suppress lymphocyte activity and induce  
PT apoptosis  
PS Claim 14; Pages 40-41; 53pp; English.  
CC The present claimed sequence represents a mouse glucocorticoid  
CC induced TNFR-family related protein variant B (G1TR-B). The invention  
CC also claims for G1TR (W49016) and G1TR-C (W49018). The G1TRs are  
CC claimed to be useful for stimulating lymphocyte activity and cell  
CC death rescue. G1TR antagonists are claimed to be useful for suppressing  
CC the lymphocyte activity and for inducing apoptotic deletion. G1TR  
CC cDNAs and the proteins they encode are also claimed to be useful for  
CC suppressing growth of tumour cells over-expressing G1TR or for the  
CC treatment of refractory hodgkin's disease.  
SQ Sequence 294 AA;

Query Match 82.5%; Score 188; DB 33; Length 294;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mgawamlygvsmclvldlgqpsvveepgpgkvgngsgnntccslyapgedckerc 60  
QY 1 MGAWAMLYGVSMCLVLDLGQPSVVEEPGPGKVGNGSGNNTCCSLYAPGEDCKERC 60  
Db 61 icvtpeyhcgdpqckickhpcqgrvesqgdvfgfrvacamgtfsagrdghcrllwt 120  
QY 61 ICVTPEYHCGDPOCKICKHPCQGRVESQGDVFGFRVACAMGTFSAGRDGHCRLLWT 120  
Db 121 ncsqfgfltmfpgnktthnavcipeplteqyghltviflvmaaciffittvqlghiwql 180  
QY 121 NCSQFGFLTMFPGNKTTHNAVCIPEPLTEQYGHLTVIFLVMAACIFFITTVQLGLHIWQL 180  
Db 181 rrqhmcr 188  
QY 181 RRQHMCR 188

RESULT 5  
ID W37840 standard; Protein; 228 AA.  
AC W37840;  
DE Truncated human 312C2 protein from clone A8 amino acid sequence.  
KW Human 312C2 T cell protein; thymus cell; spleen cell; T cell;  
KW antigen-specific T cell proliferation; cytokine production by T-cell;  
KW apoptosis; cancer; haematopoietic cells; lymphoid cell;  
KW autoimmune disorders.  
OS Homo sapiens.  
PN W09806842-A1.  
PD 19-FEB-1998.  
PF 14-AUG-1997; U13931.  
PR 07-OCT-1996; US-027901.  
PR 16-AUG-1996; US-689943.  
PA (SCHE ) SCHERING CORP.  
PI Gorman DM, Randall TD, Zlotnik A;  
DR WPI; 98-159534/14.  
DR N-PSDB; V19154.  
PT Isolated 312C2 T cell gene - used to develop products for treating,  
PT e.g. cancers, auto-immune disorders, transplantation rejection and  
PT other T cell disorders  
PS Disclosure: Pages 61-62; 71pp; English.  
CC This is the amino acid sequence of the truncated human 312C2 T cell  
CC protein from clone A8. The 312C2 proteins are expressed in thymus  
CC cells and are induced on T cells and spleen cells following activation.  
CC Engagement of 312C2 stimulates proliferation of T cell clones, and  
CC antigen-specific proliferation and cytokine production by T-cells, and  
CC potentiates T cell expansion or apoptosis. The products can be used  
CC in the treatment of conditions associated with abnormal physiology or  
CC development, including abnormal proliferation, e.g. cancerous  
CC conditions or degenerative conditions. They can be used in the  
CC regulation or development of haematopoietic cells, e.g. lymphoid cells  
CC which affect immunological responses, e.g. autoimmune disorders.  
SQ Sequence 228 AA;

PT New isolated glucocorticoid induced TNFR related polypeptide - used  
PT to stimulate lymphocyte activity and cell death rescue, useful to,  
PT e.g. develop products to suppress lymphocyte activity and induce  
PT apoptosis  
PS Claim 15; Pages 43-44; 53pp; English.  
CC The present claimed sequence represents a mouse glucocorticoid  
CC induced TNFR-family related protein variant C (G1TR-C). The invention  
CC also claims for G1TR (W49016) and G1TR-B (W49017). The G1TRs are  
CC claimed to be useful for stimulating lymphocyte activity and cell  
CC death rescue. G1TR antagonists are claimed to be useful for suppressing  
CC the lymphocyte activity and for inducing apoptotic deletion. G1TR  
CC cDNAs and the proteins they encode are also claimed to be useful for  
CC suppressing growth of tumour cells over-expressing G1TR or for the  
CC treatment of refractory hodgkin's disease.  
SQ Sequence 222 AA;

Query Match 82.5%; Score 188; DB 33; Length 222;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mgawamlygvsmclvldlgqpsvveepgpgkvgngsgnntccslyapgedckerc 60  
QY 1 MGAWAMLYGVSMCLVLDLGQPSVVEEPGPGKVGNGSGNNTCCSLYAPGEDCKERC 60  
Db 61 icvtpeyhcgdpqckickhpcqgrvesqgdvfgfrvacamgtfsagrdghcrllwt 120  
QY 61 ICVTPEYHCGDPOCKICKHPCQGRVESQGDVFGFRVACAMGTFSAGRDGHCRLLWT 120  
Db 121 ncsqfgfltmfpgnktthnavcipeplteqyghltviflvmaaciffittvqlghiwql 180  
QY 121 NCSQFGFLTMFPGNKTTHNAVCIPEPLTEQYGHLTVIFLVMAACIFFITTVQLGLHIWQL 180  
Db 181 rrqhmcr 188  
QY 181 RRQHMCR 188

RESULT 4  
ID W49017 standard; Protein; 294 AA.  
AC W49017;  
DE 29-SEP-1998 (first entry)  
DE Mouse glucocorticoid induced TNFR-family related protein variant B.  
KW Mouse glucocorticoid induced TNFR-family related protein; lymphocyte;  
KW G1TR; tumour necrosis factor; apoptosis; hodgkin's disease; G1TR-B;  
KW G1TR-C.  
OS Mus sp.  
PN W09824895-A1.  
PD 11-JUN-1998.  
PF 08-NOV-1997; E06252.  
PR 02-DEC-1996; GB-025074.  
PA (PHAA ) PHARMACIA & UPJOHN SPA.  
PI Riccardi C;  
DR WPI; 98-333315/29.  
DR N-PSDB; V32774.



DR WPI; 99-070066/06.  
 DR N-PSDB; X00674.  
 PT New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PS Claim 11; Page 307; 385pp; English.  
 CC This sequence represents a secreted human protein encoded by the gene  
 CC clone detailed in the descriptor line. The gene can be used to generate  
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc  
 CC portion (e.g. X00602) for increasing the stability of the fused protein  
 CC as compared to the human protein only.  
 CC The invention relates to 87 novel genes and their fragments (nucleic acid  
 CC sequences: X00611-X00724; amino acid sequences W67807-W68004) which  
 CC are useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. Also, pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the 87 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see X00611 for described  
 CC uses).  
 Sequence 69 AA;

Query Match 2.6%; Score 6; DB 39; Length 69;  
 Best Local Similarity 100.0%; Pred. No. 9.87e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 pcqpgq 7

QY 81 PCQPGQ 86

RESULT 9  
 ID W5491 standard; Protein; 140 AA.

AC W5491;  
 DT 30-JUN-1998 (first entry)  
 DE H. pylori ORF hp6el0967.23476509.f2.6 secreted protein.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
 KW bacterium.  
 OS Helicobacter pylori.  
 PN W09737044-AL.  
 PD 09-OCT-1997.  
 PF 27-MAR-1997; U05223.  
 PR 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 25-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 (ASTR ) ASTRA AB.  
 PI Alm RA, Smith D;  
 DR WPI; 97-503122/46.  
 DR N-PSDB; V24900.  
 PT Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claims 14,94; Pages 697-698; 1145pp; English.  
 CC This sequence is a H. pylori secreted protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The DNA and probes derived from it may be used for the  
 CC identification of H. pylori in a sample, and the diagnosis of  
 CC H. pylori infection. Nucleic acid sequences complementary to the  
 CC DNA act as antisense sequences, and can be used to prevent the  
 CC translation of H. pylori mRNA. Antibodies against the protein can  
 CC be used in immunoassays to evaluate the abundance and distribution  
 CC of H. pylori-specific antigens. The genomic sequence of H. pylori  
 CC (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were  
 CC analysed for ORF of at least 180 nucleotides, and the predicted  
 CC coding regions defined by computer evaluation. To identify likely  
 CC H. pylori antigens for vaccine development, the amino acid

CC sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having  
 CC identified and determined the sequences of interest, particular  
 CC regions can be isolated from H. pylori by PCR amplification for  
 CC recombinant polypeptide production, e.g. in E. coli hosts.  
 SQ Sequence 140 AA;

Query Match 2.6%; Score 6; DB 29; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 9.87e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 57 lsaeda 62

QY 198 LSAEDA 203

RESULT 10  
 ID W5326 standard; Protein; 165 AA.

AC W5326;  
 DT 15-JUN-1998 (first entry)  
 DE H. pylori ORF hp2p10625orf5 protein.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 OS Helicobacter pylori.  
 PN W09737044-AL.  
 PD 09-OCT-1997.  
 PF 27-MAR-1997; U05223.  
 PR 06-DEC-1996; US-761318.  
 PR 29-MAR-1996; US-625811.  
 PR 02-APR-1996; US-758731.  
 PR 25-OCT-1996; US-736905.  
 PR 28-OCT-1996; US-738859.  
 (ASTR ) ASTRA AB.  
 PI Alm RA, Smith D;  
 DR WPI; 97-503122/46.  
 DR N-PSDB; V24735.  
 PT Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PS Claim 14; Pages 552-553; 1145pp; English.  
 CC This sequence is a H. pylori protein of unspecified function.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 SQ Sequence 165 AA;

Query Match 2.6%; Score 6; DB 29; Length 165;  
 Best Local Similarity 100.0%; Pred. No. 9.87e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 82 lsaeda 87

QY 198 LSAEDA 203

RESULT 11

ID R29897 standard; Protein; 167 AA.  
AC R29897;  
DT 26-APR-1993 (first entry)  
DE HCV NS4-NS5 peptide N29-3.  
KW Clone; polypeptide; NS4-NS5; Hepatitis C; Virus; HCV; serum; HC;  
KW transcriptase; cDNA; primer; allele; core; region; upstream;  
KW hydrophilic; turn structure; alpha helix; beta sheet; antigen;  
KW determinant; antiserum.  
OS Hepatitis C virus  
PN EP-518313-A.  
PD 16-DEC-1992.  
PF 11-JUN-1992; 109812.  
PR 11-JUN-1991; JP-139268.  
PR 12-JUL-1991; JP-172794.  
PR 07-OCT-1991; JP-287008.  
PR 16-DEC-1991; JP-332329.  
PR 20-APR-1992; JP-099957.  
PA (MITU) MITSUBISHI KASEI CORP.  
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;  
PI Teranishi Y;  
PI N-PSDB; 92-417213/51.  
PT New hepatitis C virus gene and its encoded protein - used for  
PT diagnosing and vaccinating against hepatitis C virus infections  
PS Disclosure; Page 244-45; 305pp; English.  
CC The sequences given in R29871-906 and R29533 are encoded by various  
CC clones of the NS4-NS5 regions of the Hepatitis C virus (HCV) gene of  
CC the invention. These NS4-NS5 RNA sequences were isolated from the  
CC serum of a patient suffering from hepatitis C (HC). The isolated RNA  
CC sequences were converted into cDNA using transcriptase in the presence  
CC of one of the primer sequences given in Q32565-77. The sequences were  
CC then amplified using primer pairs. The cDNA sequences isolated  
CC represent different alleles of the same region of the HCV gene.  
CC Sequence analysis shows that these clones represent the core region  
CC and some upstream sequences of HCV. These polypeptides are thought to  
CC contain a highly hydrophilic region which can adopt a "turn structure"  
CC which is not an alpha helix or a beta sheet. These polypeptides are  
CC thought to act as antigen determinants and are highly reactive with  
CC antiserum raised against HCV-associated antigens. See also Q32436.  
SQ Sequence 167 AA;  
Query Match 2.6%; Score 6; DB 6; Length 167;  
Best Local Similarity 100.0%; Pred. No. 9.87e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 134 pgqrve 139  
QY 84 PGQVRVE 89  
RESULT 12  
ID R29896 standard; Protein; 171 AA.  
AC R29896;  
DT 26-APR-1993 (first entry)  
DE HCV NS4-NS5 peptide N29-2.  
KW Clone; polypeptide; NS4-NS5; Hepatitis C; Virus; HCV; serum; HC;  
KW transcriptase; cDNA; primer; allele; core; region; upstream;  
KW hydrophilic; turn structure; alpha helix; beta sheet; antigen;  
KW determinant; antiserum.  
OS Hepatitis C virus  
PN EP-518313-A.  
PD 16-DEC-1992.  
PF 11-JUN-1992; 109812.  
PR 11-JUN-1991; JP-139268.  
PR 12-JUL-1991; JP-172794.  
PR 07-OCT-1991; JP-287008.  
PR 16-DEC-1991; JP-332329.  
PR 20-APR-1992; JP-099957.  
PA (MITU) MITSUBISHI KASEI CORP.  
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;  
PI Teranishi Y;  
PI N-PSDB; 92-417213/51.  
PT New hepatitis C virus gene and its encoded protein - used for  
PT diagnosing and vaccinating against hepatitis C virus infections  
PS Disclosure; Page 244-45; 305pp; English.  
CC The sequences given in R29871-906 and R29533 are encoded by various  
CC clones of the NS4-NS5 regions of the Hepatitis C virus (HCV) gene of  
CC the invention. These NS4-NS5 RNA sequences were isolated from the  
CC serum of a patient suffering from hepatitis C (HC). The isolated RNA  
CC sequences were converted into cDNA using transcriptase in the presence  
CC of one of the primer sequences given in Q32565-77. The sequences were  
CC then amplified using primer pairs. The cDNA sequences isolated  
CC represent different alleles of the same region of the HCV gene.  
CC Sequence analysis shows that these clones represent the core region  
CC and some upstream sequences of HCV. These polypeptides are thought to  
CC contain a highly hydrophilic region which can adopt a "turn structure"  
CC which is not an alpha helix or a beta sheet. These polypeptides are  
CC thought to act as antigen determinants and are highly reactive with  
CC antiserum raised against HCV-associated antigens. See also Q32436.  
SQ Sequence 167 AA;  
Query Match 2.6%; Score 6; DB 6; Length 167;  
Best Local Similarity 100.0%; Pred. No. 9.87e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 134 pgqrve 139  
QY 84 PGQVRVE 89  
RESULT 13  
ID P90160 standard; Protein; 171 AA.  
AC P90160;  
DT 1-NOV-1989 (first entry)  
DE Sequence of hepatitis C virus cDNA insert in clone 35f  
KW Hepatitis C virus; clone 35f; clone 39c; probe; vaccine.  
OS Pan troglodytes  
FH Key Location/Qualifiers  
FT region 1..14  
FN GB2212511-A.  
PD 26-JUL-1989.  
PF 18-NOV-1988; G27024.  
PR 18-NOV-1987; US-122714.  
PA (CHIR) Chiron Corporation.  
PI Houghton M, Choo QL, Kuo G;  
PI WPI; 89-215054/30.  
DR N-PSDB; N90329.  
PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,  
PT polypeptide(s) and antibodies for diagnosis, prevention and  
PT treatment of infection.  
PS Disclosure; fig 28; 235pp; English.  
CC The sequence is the peptide encoded by the hepatitis C virus  
CC (HCV) cDNA insert in clone 35f (see N90329). The polypeptides  
CC are used to diagnose HCV-induced NANBH, to raise antibodies for  
CC immunoassay or treatment, or to produce vaccines.  
CC The region shown overlaps with clone 39c.  
SQ Sequence 171 AA;  
Query Match 2.6%; Score 6; DB 1; Length 171;  
Best Local Similarity 100.0%; Pred. No. 9.87e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 130 pgqrve 135  
QY 84 PGQVRVE 89  
RESULT 14  
ID P92043 standard; Protein; 171 AA  
AC P92043  
DT 02-MAR-1990 (first entry)  
DE Sequence encoded in the hepatitis C virus (HCV) cDNA insert in clone 35f.  
KW Hepatitis C virus (HCV); non-A, non-B hepatitis (HNBH).  
OS Hepatitis C virus.  
FH Key Location/Qualifiers

FT region 1..14  
FT EP-318216-A. /\*tag= a  
PD 31-MAY-1989.  
PF 18-NOV-1988; 310922.  
PR 14-NOV-1988; US-271450; US-122714.  
PA (CHIR) Chiron Corp.  
PI Houghton M, Choo Q-L, Kuo G;  
DR WPI: 89-159274/22.  
DR N-PSDB: N92099.  
PT Purified hepatitis C virus  
PT - and associated nucleic acids and polypeptide(s)  
PS Claim 13; Figure 28; 139pp; English.  
CC It is the sequence encoded in the hepatitis C virus (HCV) cDNA insert in  
CC clone 35f. Tag a = the region of overlap with the HCV antigen encoded in  
CC clone 39c. It is antigenic and could be used in immunoassay reagents  
CC and vaccines and to generate antibodies useful in diagnosis and passive  
CC immunotherapy for HCV infection/non-A, non-B hepatitis.  
SQ Sequence 171 AA;  
Query Match 2.6%; Score 6; DB 1; Length 171;  
Best Local Similarity 100.0%; Pred.No. 9.87e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 130 pggrve 135  
QY 84 PGORVE 89  
|||||  
RESULTS 15  
ID R29874 standard; Protein: 171 AA.  
AC R29874.  
DT 26-APR-1993 (first entry)  
DE HCV NS4-NS5 peptide N29-1, N29-2, N29-3.  
KW Clone; polypeptide; NS4-NS5; Hepatitis C; Virus; HCV; serum; HC;  
KW transcriptase; cDNA; primer; allele; core; region; upstream;  
KW hydrophilic; turn structure; alpha helix; beta sheet; antigen;  
KW determinant; antiserum.  
OS Hepatitis C virus  
FH Key Location/Qualifiers  
FT misc\_difference 61  
FT /label= Glu, Lys  
FT misc\_difference 155  
FT /label= Ala, Ser  
PN EP-518313-A.  
PD 16-DEC-1992.  
PR 11-JUN-1992; 109812.  
PR 11-JUN-1991; JP-139268.  
PR 12-JUL-1991; JP-172794.  
PR 07-OCT-1991; JP-287008.  
PR 16-DEC-1991; JP-332329.  
PR 20-APR-1992; JP-099957.  
PA (MITU ) MITSUBISHI KASEI CORP.  
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;  
PI Teranishi Y;  
DR WPI: 92-417213/51.  
DR N-PSDB: Q32505.  
PT New hepatitis C virus gene and its encoded protein - used for  
PT diagnosing and vaccinating against hepatitis C virus infections  
PS Disclosure; Page 197-98; 305pp; English.  
CC The sequences given in R29871-906 and R29533 are encoded by various  
CC clones of the NS4-NS5 regions of the Hepatitis C Virus (HCV) gene of  
CC the invention. These NS4-NS5 RNA sequences were isolated from the  
CC serum of a patient suffering from hepatitis C (HC). The isolated RNA  
CC sequences were converted into cDNA using transcriptase in the presence  
CC of one of the primer sequences given in Q32565-77. The sequences were  
CC then amplified using primer pairs. The cDNA sequences isolated  
CC represent different alleles of the same region of the HCV gene.  
CC Sequence analysis shows that these clones represent the core region  
CC and some upstream sequences of HCV. These polypeptides are thought to  
CC contain a highly hydrophilic region which can adopt a "turn structure"  
CC which is not an alpha helix or a beta sheet. These polypeptides are  
CC thought to act as antigen determinants and are highly reactive with

CC antiserum raised against HCV-associated antigens. See also Q32436.  
SQ Sequence 171 AA;  
Query Match 2.6%; Score 6; DB 6; Length 171;  
Best Local Similarity 100.0%; Pred.No. 9.87e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 134 pggrve 139  
QY 84 PGORVE 89  
|||||  
Search completed: Mon Jul 12 12:08:40 1999  
Job time : 29 secs.